WO 01/75180 PCT/US01/09392

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 October 2001 (11.10.2001)

PCT

(10) International Publication Number
WO 01/75180 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US01/09392

(22) International Filing Date: 23 March 2001 (23.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/538,709 30 March 2000 (30.03.2000) US

(71) Applicant *(for all designated States except US)*: **QBI ENTERPRISES LTD.** [IL/IL]; Weizmann Scientific Park, P.O. Box 4071, Ness Ziona (IL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **ULANOVSKY, Levy** [US/US]; 850 Matadero Avenue, Palo Alto, CA 94306-2605 (US). **MUGASIMANGALAM, Raja** [IN/IN]; 105 3rd Main, CIL Layout, Sanjaynagar, Bangalor 560094 (IN). **EINAT, Paz** [IL/IL]; Teper 3/27, 74042 Ness-Ziona (IL). **ZEZIN-SONKIN, Dina** [IL/IL]; Ezra Street 25/6, 76201 Rehovot (IL). **SHLOMIT, Gilad** [IL/IL]; Tel Nof, 70700 Gedera (IL).

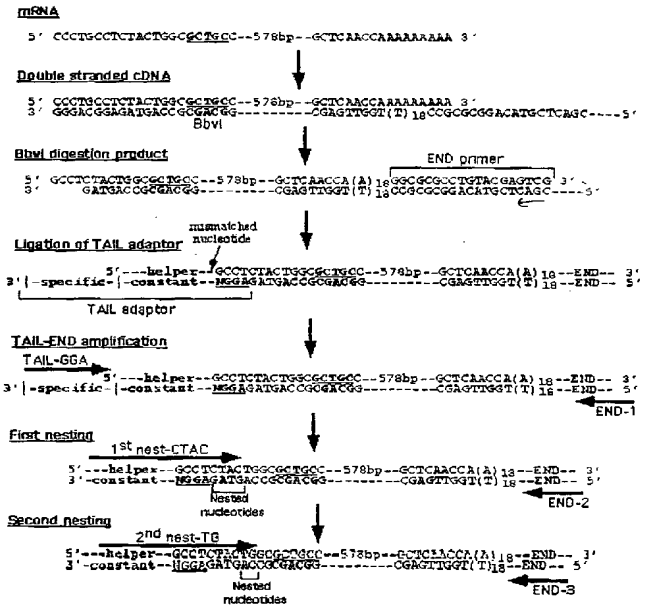
(74) Agents: **KOWALSKI, Thomas, J.** et al.; Frommer Lawrence & Haug LLP, 745 Fifth Avenue, New York, NY 10151 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,

[Continued on next page]

(54) Title: SEQUENCE-DEPENDENT GENE SORTING TECHNIQUES



(57) Abstract: This invention provides a method of sorting genes comprising: (1) preparing ds cDNA molecules from mRNA molecules; (2) digesting the ds cDNA molecules; (3) ligating to the digested cDNA molecules a set of dsDNA oligonucleotide adaptors; (4) amplifying the ligated cDNA molecules; and (5) sorting the amplified cDNA molecules into non-redundant groups. This invention also provides two additional methods of sorting genes. This invention further provides a method of making sub-libraries of ligation sets and a method of making sub-libraries of genetic vectors.

WO 01/75180 A2

WO 01/75180 A2

IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TITLE

SEQUENCE-DEPENDENT GENE SORTING TECHNIQUES

FIELD OF THE INVENTION

The present invention discloses techniques for simply and efficiently sorting expressed genes into non-redundant groups of cDNA molecules reverse-transcribed from any source of eukaryotic RNA. These groups of cDNA molecules can themselves be used for genetic analyses according to methods in the art, or they can be further sorted according to the techniques of the present invention. By applying these techniques one can obtain a collection of non-redundant subgroups of cDNA molecules, with every expressed-gene transcript from the original mRNA sample uniquely represented in its own subgroup. The method further provides a stage in which each expressed-gene transcript is found in one tube, i.e. "one gene per well." Uses of the present invention include isolation, identification and analysis of genes, analysis and diagnosis of disease states, study of cellular differentiation, and gene therapy.

BACKGROUND OF THE INVENTION

The production of cDNA or gene libraries has involved cloning by the use of cloning vectors placed in host organisms such as bacteria or yeast. These libraries suffer from redundancy: they contain either multiple copies of particular cDNA sequences, or multiple cDNA fragments from each expressed gene, or both. This redundancy persists in all of current normalization procedures. The presence in a collection of cDNAs of multiple copies of particular cDNA sequences, and/or multiple cDNA fragments from each expressed gene, can result in pointless duplication of research efforts and other significant inefficiencies.

U.S. Patent No. 5,707,807 concerns the creation of subgroups of DNA by repeated digestions with a number of restriction enzymes, followed by ligation with adaptors having a common primer template, PCR amplification and, finally, comparison of patterns of PCR products separated by polyacrylamide-gel electrophoresis. The method of this patent creates groups of DNA molecules. However, because each PCR step indiscriminately amplifies all ligated DNA molecules in each sample, the method has a limited capacity to sort DNA into non-redundant groups.

Unrau and Deugau (1994) Gene 145:163-169 concerns characterizing fragments of digested DNA by the sequences of their cohesive ends and their lengths, optionally aided by PCR. However, each PCR step indiscriminately amplifies all ligated DNA molecules in each sample, and amplifies numerous DNA fragments per gene. The method does not yield non-redundant groups of genes.

U.S. Patent No. 5,728,524 concerns obtaining groups of DNA molecules by using pools of adaptors ligated to digested DNA, followed by PCR. Each PCR step amplifies numerous DNA fragments per gene. The method fails to produce non-redundant groups of genes.

Smith concerns a general method for PCR amplification of type II restriction fragments by ligation of adaptors with degenerate end sequences complementary to cohesive ends of digested DNA fragments. Each PCR step amplifies numerous DNA fragments per gene. The method fails to produce non-redundant gene groups.

U.S. Patent No. 5,871,697 concerns classifying DNA sequences by making extensive use of comparative databases and fragment-length and restriction-digest information. The patent concerns DNA digestion and ligation of adaptors with priming sequences specific for a particular restriction enzyme. The method in this patent does not aim at the production of non-redundant groups of genes.

Throughout this application, various references are cited author by and publication date. Each of these publications and each of the documents cited in each of these publications, and each document referenced or cited in the publication cited documents are hereby incorporated herein by reference.

OBJECTS AND SUMMARY OF THE INVENTION

The present invention provides novel methods for producing a non-redundant cDNA or gene library. The methods sort DNA on a sequence-dependent basis into non-redundant groups. At the same time, however, these methods eliminate the need to determine any of the DNA sequences prior to sorting and identifying genes.

One object of the present invention to provide a method of sorting cDNA or genes into non-redundant groups, which can then be analyzed by techniques known the art. One of many such techniques is the cDNA microarray method in which the cDNA clones derived from the present invention are used to produce the array that is then examined by hybridization to determine differential gene expression.

Another technique is differential display of gel-electrophoresis patterns involving mRNA sources to analyze biological models such as disease states or cellular differentiation. In application of this technique the groups derived from the present invention can be used for differential display of gel-electrophoresis patterns.

Another object of the present invention is providing a method of obtaining a collection of non-redundant subgroups of cDNA molecules, with every expressed-gene transcript from an original mRNA sample uniquely represented in its own subgroup, i.e. "one gene per well." Such isolated genes have a wide-variety of uses, notably including gene therapy and analysis of the human genome.

The present invention provides a method of sorting genes and/or gene fragments comprising the following steps (herein called "Method I"):

(1) preparing ds cDNA molecules from mRNA molecules by reverse transcription, using a poly-T primer optionally having a general primer-template sequence upstream from the poly-T sequence, yielding ds cDNA molecules having the poly-T sequence, optionally having the general primer-template sequence;

(2) digesting the ds cDNA molecules with a restriction enzyme that produces digested cDNA molecules with cohesive ends having overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(3) ligating to the digested cDNA molecules a set of dsDNA oligonucleotide adaptors, each of which adaptor has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible overhanging ssDNA sequences of the digested cDNA, at the opposite end a specific primer-template sequence specific for the ssDNA adaptor complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set;

(4) amplifying by separate polymerase chain reactions (PCRs) the ligated cDNA molecules, utilizing for each separate PCR a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences; and

(5) sorting the amplified cDNA molecules into non-redundant groups by collecting the amplification products after each separate PCR, each group of amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the PCR.

One embodiment of the present invention according to the principles of Method I, comprises a complete set of oligonucleotide adaptors and specific primers, containing an oligonucleotide adaptor and a specific primer complementary to each of the possible overhanging ssDNA sequences of the digested cDNA.

Another embodiment of the present invention according to the principles of Method I further comprises:

(6) amplifying the sorted, non-redundant groups of cDNA molecules by nesting PCR, each amplification utilizing a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template sequence, as well as one of a set of nesting primers with the following general formula:

5'-|sequence complementary to the constant sequence of the oligonucleotide adaptors|-NI_x-|1-5 nucleotides complementary to one of the possible sequences of 1-5 nucleotides immediately upstream from the overhanging ssDNA sequence on the cDNA|-3' where N is an arbitrary nucleotide; I is inosine; and x=1,2,3 or 4, being one fewer than the constant number of nucleotides in the overhanging ssDNA sequences; and

(7) sorting the amplified cDNA molecules into non-redundant subgroups by collecting the amplification products after each separate nesting PCR, each non-redundant subgroup of cDNA molecules determined by the particular nested primer that complemented the 1-5 nucleotides immediately upstream from the overhanging ssDNA sequence on the cDNA.

Another embodiment of the present invention according to the principles of Method I further comprises conducting further PCRs with further nesting primers complementary to the next immediately upstream cDNA nucleotides, thereby sorting the amplified cDNA molecules further into non-redundant subgroups.

A preferred embodiment according to the principles of Method I further comprises conducting further PCRs with further nesting primers complementary to the next immediately upstream cDNA nucleotides until each non-redundant subgroup contains only one type of cDNA molecule, with every expressed-gene transcript in the mRNA sample uniquely represented in one of the non-redundant subgroups.

The present invention also concerns a method of sorting genes and/or gene fragments comprising the following steps (herein called "Method II"):

(1) preparing ds cDNA molecules from mRNA molecules by reverse transcription, using a poly-T primer optionally having a general primer-template sequence upstream from the poly-T sequence, yielding ds cDNA molecules having the poly-T sequence, optionally having the general primer-template sequence;

(2) digesting the ds cDNA molecules with a first restriction enzyme that produces digested cDNA molecules with cohesive ends having first overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(3) ligating to the digested cDNA molecules a set of dsDNA oligonucleotide adaptors, each of which adaptor has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible first overhanging ssDNA sequences of the digested cDNA, at the opposite end a specific primer-template sequence specific for the ssDNA adaptor complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set, and that contains a recognition site for a second restriction enzyme that can cleave the ligated cDNA molecules at a point further from the ligated oligonucleotide adaptor than the overhanging ssDNA sequences of the digested cDNA, and can create cohesive ends having second overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(4) amplifying by separate PCRs, the ligated cDNA molecules, utilizing for each separate PCR a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences; and

(5) sorting the amplified cDNA molecules into non-redundant groups by collecting the amplification products after each separate PCR, each group of amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the PCR.

One embodiment of the present invention according to the principles of Method II comprises using a complete set of oligonucleotide adaptors and specific primers, containing an oligonucleotide adaptor and a specific primer complementary to each of the possible first overhanging ssDNA sequences of the digested cDNA.

Another embodiment of the present invention according to the principles of Method II further comprises

(6) digesting the sorted non-redundant groups of cDNA molecules with the second restriction enzyme, cleaving the ligated cDNA molecules at a point further from the ligated oligonucleotide adaptor than the overhanging ssDNA sequences of the digested cDNA, and creating cohesive ends having second overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(7) ligating to the digested cDNA molecules a set of nesting dsDNA oligonucleotide adaptors, each of which adaptor has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible second overhanging ssDNA sequences of the digested cDNA, at the opposite end a specific primer-template sequence unique for the ssDNA adaptor complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set, and that contains the recognition site for the second restriction enzyme;

(8) amplifying by separate PCRs, the ligated cDNA molecules, utilizing for each separate PCR a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences; and

(9) sorting the amplified cDNA molecules into non-redundant subgroups by collecting the amplification products after each separate PCR, each

subgroup of amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the PCR.

One embodiment of the present invention according to the principles of Method II further comprises using a complete set of nesting dsDNA oligonucleotide adaptors, containing an oligonucleotide adaptor complementary to each of the possible second overhanging ssDNA sequences of the digested cDNA.

Another embodiment according to the principles of Method II further comprises conducting further PCRs using further nesting oligonucleotide adaptors, optionally with different restriction enzymes and recognition sites, thereby sorting the amplified cDNA molecules further into non-redundant subgroups.

A preferred embodiment according to the principles of Method II further comprises conducting further ligations with further nesting oligonucleotide adaptors, optionally with different restriction enzymes and recognition sites, until each non-redundant subgroup contains only one type of cDNA molecule, with every expressed-gene transcript in the mRNA sample uniquely represented in one of the non-redundant subgroups.

The present invention also provides a method (Method III) of sorting genes and/or gene fragments comprising the steps of:

(1) preparing ds cDNA molecules from mRNA molecules by reverse transcription, using a poly-T primer having a general primer-template sequence upstream from the poly-T sequence that includes a recognition sequence for a restriction enzyme, yielding ds cDNA molecules having the poly-T sequence, having the general primer-template sequence;

(2) dividing the cDNA into N pools, wherein N is 1 to 25, by digesting the ds cDNA molecules with different restriction enzymes that produce digested cDNA molecules with cohesive ends having overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(3) ligating to the digested cDNA molecules of each pool a set of dsDNA oligonucleotide adaptors, each of which adaptor has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible overhanging ssDNA sequences of the digested cDNA, at the opposite end a specific primer-template sequence specific for the ssDNA adaptor

complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set;

(4) amplifying by separate PCRs the ligated cDNA molecules of each pool, utilizing for each separate PCR a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences;

(5) sorting the amplified cDNA molecules from each pool into non-overlapping groups by collecting the amplification products after each separate PCR, each group of amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the PCR, wherein each of the restriction enzymes digests the N separate cDNA pools into 64 or 256 non-redundant sub-groups; and

(6) digesting cDNA fragments in each non-redundant sub-group of the cDNA pools with different restriction enzymes and further purifying the digested cDNA fragments by removing the small end fragments produced by the digestion.

This invention also provides a method of making sub-libraries of ligation sets by ligating restriction enzyme digested fragments generated by method III into a plasmid vector that have recognition sequence for said restriction enzymes and predigesting with these enzymes to make 64xN or 256xN sets of ligations, wherein N is 1 to 25.

This invention further provides a method of making sub-libraries of bacterial colonies, wherein the set of ligations, generated in the method of making sub-libraries of ligation sets, are transformed into bacteria and plated onto bacterial growth plates to produce bacteria colonies containing each of the 64xN or 256xN non-redundant subgroups of cDNA fragments, wherein N is 1 to 25.

In one embodiment of method III, N is two and the restriction enzyme in step (1) comprises *Asc*I or another similar rare restriction enzyme.

BRIEF DESCRIPTION OF THE DRAWINGS

The following Detailed Description, given by way of example, but not intended to limit the invention to specific embodiments described, may be

understood in conjunction with the accompanying Figures, incorporated herein by reference, in which:

Figure 1 schematically illustrates one embodiment, Primer Nesting Option, of the principles of Method I showing a flow chart using a specific sequence as an example (SEQ ID NOs: 36-48).

Figure 2 schematically illustrates one embodiment, Ligation Nesting Option [64-256-16], of the principles of Method II showing a flow chart using a specific sequence as an example. The first four steps are shown in Figure 1 (SEQ ID NOs: 49-62).

Figure 3 schematically illustrates an alternative embodiment, Ligation Nesting Option [64-64-64], of the principles of Method II showing a flow chart using a specific sequence as an example. The first four steps are shown in Figure 2 (SEQ ID NOs: 63-76).

Figure 4 shows ligation specificity permitting isolation of the rat albumin gene using A. standard ligation conditions and B. the methods of the present invention.

Figure 5 shows ligation specificity using human GAPDH gene with a particular set of ligation adaptors using the methods of the present invention. The results are shown as GAPDH-rev PCR analysis of GAPDH ligation specificity.

Figure 6 shows PCR amplification products derived from Jurkat-cell mRNA using a particular set of ligation adaptors according to the methods of the present invention. The double stranded cDNA (derived from Jurkat cells) that was ligated to the mix of all 64 "Tail adaptor set 1" adaptors was used as template. The cDNA group ligated to each adaptor was amplified separately using the specific Tail primer and the END primer. The figure shows the products of all 64 Tail-END amplification reactions. Amplification products were separated on a 1.5% agarose gel and ethidium bromide staining was used to visualize the DNA.

Figure 7 is a Southern blot of PCR amplification products derived from Jurkat-cell mRNA showing ligation specificity according to the methods of the present invention. The agarose gel shown in Figure 2 was blotted onto nylon membrane (Nytran, Schleicher & Schuell). The membrane was then hybridized with a radioactive (^{32}P) probe specific to the human GAPDH gene. The specific signal

was obtained in the correct "AGG" lane only. A weaker signal, observed in the "CCC" lane is not of the correct size and can be caused by spurious amplification of the abundant GAPDH cDNA by the "CCC" Tail primer alone.

Figure 8 shows isolation of three different genes obtained by using the first nesting PCR primers according to the methods of the present invention. The specific END-Tail groups, expected to contain the GAPDH, KU autoantigen and fibrillarin cDNAs, were used as a template for nesting PCR. Nesting primers from the "1st Nest 256", expected to amplify these three genes were used. Amplification products were separated on a 1.5% agarose gel and ethidium bromide staining was used to visualize the DNA. For the GAPDH and KU antigen cDNAs single bands of the correct size are observed. For fibrillarin cDNA three bands are observed, one of them, the middle 650bp band, is of the expected size.

Figure 9 shows isolation of three different genes obtained by using second nesting primers according to the methods of the present invention. Products of the 1st nesting reactions were used as template for the second nesting. Primers from the "2nd nest 16" set were chosen that are expected to amplify the three cDNAs. As expected, single strong bands were obtained for GAPDH and KU autoantigen cDNAs. For fibrillarin, the second nesting step separated the three bands and only the correct 650bp band was obtained.

Figure 10 shows the general structure of the primer and the primer set J2 (SEQ ID NOs: 77- 141).

Figure 11 shows tail primers set J2 (SEQ ID NOs:142-205).

Figure 12 shows tail primers (set number 2) (SEQ ID NOs: 206-265)

Figure 13 shows tail primers set 256 (SEQ ID NOs: 266-521).

Figure 14 shows first nesting primers 256 for tail adaptor 64 set 1 (SEQ ID NOs: 522-777).

Figure 15 shows first nesting primers 64 for tail adaptor 64 set 1 (SEQ ID NOs: 778-841).

Figure 16 shows second nesting primers 64 for tail adaptor 64 set 1 (SEQ ID NOs: 842-905).

Figure 17 shows second nesting primers 16 for tail adaptor 64 set 1 (SEQ ID NOs: 906-921).

Figure 18 shows tail adaptors set 256 (SEQ ID NOs: 922-1177).

Figure 19 shows tail adaptors 64 (set number 1(SEQ ID NOs: 1178-1241)) and helper oligonucleotides (SEQ ID NOs: 1142-1144).

Figure 20 shows tail adaptors 64 (set number 2 (SEQ ID NOs: 1245-1308) and helper oligonucleotides (SEQ ID NOs: 1309-1311).

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides techniques for obtaining groups of non-redundant cDNA molecules, including cDNA libraries containing "one gene per well" for every gene transcript present in an original mRNA source. These techniques sort DNA on a sequence-dependent basis into non-redundant groups, using PCR combined with (1) an initial step of "differential ligation" using a pool of dsDNA ligation adaptors, each of which has an arbitrary ssDNA end and a primer template specific for the ssDNA end, and optional further steps using (2) either nesting primers (in Method I) or nesting ligation adaptors (in Method II).

Method I broadly concerns a method of sorting genes and/or gene fragments comprising the following steps:

(1) preparing ds cDNA molecules from mRNA molecules by reverse transcription, using a poly-T primer optionally having a general primer-template sequence upstream from the poly-T sequence, yielding ds cDNA molecules having the poly-T sequence, optionally having the general primer-template sequence;

(2) digesting the ds cDNA molecules with a restriction enzyme that produces digested cDNA molecules with cohesive ends having overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(3) ligating to the digested cDNA molecules a set of dsDNA oligonucleotide adaptors, each of which adaptor has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible overhanging ssDNA sequences of the digested cDNA, at the opposite end a specific primer-template sequence specific for the ssDNA adaptor complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set;

(4) amplifying by separate PCRs the ligated cDNA molecules, utilizing for each separate PCR a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences; and

(5) sorting the amplified cDNA molecules into non-redundant groups by collecting the amplification products after each separate PCR, each group of amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the PCR.

The restriction enzyme can be any enzyme that produces digested cDNA molecules with cohesive ends having overhanging ssDNA sequences of a constant number of arbitrary nucleotides. Such restriction enzymes include type IIs restriction enzymes, including *Bbv*I, *Bsp*MI, *Fok*I, *Hga*I, *Mbo*I and *Sfa*NI. Suitable type II restriction enzymes include *Bgl*I, *Bst*XI and *Sfi*I.

The groups of cDNA molecules produced by the techniques of Method I are non-redundant: only one DNA sequence will be present for each gene, since for each gene only the poly-T-containing fragment—possibly the entire gene—is primed and amplified. As used in this invention, all genes present as transcripts in a mRNA sample were obtained using complete sets of redundant adaptors. Thus, one embodiment according to the principles of Method I comprises using a complete set of oligonucleotide adaptors and specific primers, containing an oligonucleotide adaptor and a specific primer complementary to each of the possible overhanging ssDNA sequences of the digested cDNA. If the constant number of arbitrary nucleotides in the overhanging ssDNA is 3, then a complete set of adaptors includes 4^3 or 64 different oligonucleotide adaptors. If the constant number of arbitrary nucleotides is 4, then a complete set includes 4^4 or 256 different adaptors.

Another embodiment of Method I utilizes adaptors with the 3'-most nucleotide of the ssDNA complementary sequence of the oligonucleotide adaptor an arbitrary nucleotide N, which pairs with the 5'-most nucleotide of each of the possible overhanging ssDNA sequences of the digested cDNA. A complete set of this kind of adaptors contains an oligonucleotide adaptor (for a specific primer) complementary to each of the possible overhanging ssDNA sequences of the digested cDNA excluding

the 5'-most nucleotide that pairs with the arbitrary nucleotide N of the oligonucleotide adaptor.

One embodiment of the principles of Method I further comprises additional steps:

(6) amplifying the sorted non-redundant groups of cDNA molecules by nesting PCR, each amplification utilizing a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template sequence, as well as one of a set of nesting primers with the following general formula

5'-|sequence complementary to the constant sequence of the oligonucleotide adaptors|-NI_x-|1-5 nucleotides complementary to one of the possible sequences of 1-5 nucleotides immediately upstream from the overhanging ssDNA sequence on the cDNA|-3' where N is an arbitrary nucleotide; I is inosine; and x=1,2,3 or 4, being one fewer than the constant number of nucleotides in the overhanging ssDNA sequences; and

(7) sorting the amplified cDNA molecules into non-redundant subgroups by collecting the amplification products after each separate nesting PCR, each non-redundant subgroup of cDNA molecules determined by the particular nested primer that complemented the 1-5 nucleotides immediately upstream from the overhanging ssDNA sequence on the cDNA.

As before, a complete set of nesting primers can be used, which set contains a nesting primer complementary to each of the possible sequences of 1-5 nucleotides immediately upstream from the overhanging ssDNA sequence on the cDNA.

The principles of Method I can be used to conduct further PCRs with further nesting primers complementary to the next immediately upstream cDNA nucleotides, thereby sorting the amplified cDNA molecules further into non-redundant subgroups. A preferred embodiment involves conducting further PCRs with further nesting primers complementary to the next immediately upstream cDNA nucleotides until each non-redundant subgroup contains only one type of cDNA molecule, with every

expressed-gene transcript in the mRNA sample uniquely represented in one of the non-redundant subgroups, i.e. "one gene per well."

Method II broadly concerns a method of sorting genes and/or gene fragments comprising the following steps:

(1) preparing ds cDNA molecules from mRNA molecules by reverse transcription, using a poly-T primer optionally having a general primer-template sequence upstream from the poly-T sequence, yielding ds cDNA molecules having the poly-T sequence, optionally having the general primer-template sequence;

(2) digesting the ds cDNA molecules with a first restriction enzyme that produces digested cDNA molecules with cohesive ends having first overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(3) ligating to the digested cDNA molecules a set of dsDNA oligonucleotide adaptors, each of which adaptor has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible first overhanging ssDNA sequences of the digested cDNA, at the opposite end a specific primer-template sequence specific for the ssDNA adaptor complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set, and that contains a recognition site for a second restriction enzyme that can cleave the ligated cDNA molecules at a point further from the ligated oligonucleotide adaptor than the overhanging ssDNA sequences of the digested cDNA, and can create cohesive ends having second overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(4) amplifying by separate PCRs the ligated cDNA molecules, utilizing for each separate PCR a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences; and

(5) sorting the amplified cDNA molecules into non-redundant groups by collecting the amplification products after each separate PCR, each group of

amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the PCR.

The first restriction enzyme can be any that produces digested cDNA molecules with cohesive ends having overhanging ssDNA sequences of a constant number of arbitrary nucleotides. Such restriction enzymes include type IIs restriction enzymes, including *Bbv*I, *Bsp*MI, *Fok*I, *Hga*I, *Mbo*I and *Sfa*NI. Suitable type II restriction enzymes include *Bgl*I, *Bst*XI and *Sfi*I. The second restriction enzyme can be a type II restriction enzyme that cleaves the ligated cDNA molecules at a point further from the ligated oligonucleotide adaptor than the overhanging ssDNA sequences of the digested cDNA, and creates cohesive ends having second overhanging ssDNA sequences of a constant number of arbitrary nucleotides. Examples of suitable type IIs restriction enzymes include *Bsp*MI.

As in Method I, in Method II a complete set of oligonucleotide adaptors and specific primers contains an oligonucleotide adaptor and a specific primer complementary to each of the possible first overhanging ssDNA sequences of the digested cDNA. Where the 3'-most nucleotide of the ssDNA complementary sequence of the oligonucleotide adaptor is an arbitrary nucleotide N, which pairs with the 5'-most nucleotide of each of the possible first overhanging ssDNA sequences of the digested cDNA, a complete set of oligonucleotide adaptors and specific primers contains an oligonucleotide adaptor and a specific primer complementary to each of the possible first overhanging ssDNA sequences of the digested cDNA excluding the 5'-most nucleotide that pairs with the arbitrary nucleotide N of the oligonucleotide adaptor.

One embodiment of the principles of Method II further comprises additional steps:

(6) digesting the sorted non-redundant groups of cDNA molecules with the second restriction enzyme, cleaving the ligated cDNA molecules at a point further from the ligated oligonucleotide adaptor than the overhanging ssDNA sequences of the digested cDNA, and creating cohesive ends having second overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(7) ligating to the digested cDNA molecules a set of nesting dsDNA oligonucleotide adaptors, each of which adaptor has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible second overhanging ssDNA sequences of the digested cDNA, at the opposite end a specific primer-template sequence unique for the ssDNA adaptor complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set, and that contains the recognition site for the second restriction enzyme;

(8) amplifying by separate PCRs the ligated cDNA molecules, utilizing for each separate PCR a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences; and

(9) sorting the amplified cDNA molecules into non-redundant subgroups by collecting the amplification products after each separate PCR, each subgroup of amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the PCR.

A complete set of nesting dsDNA oligonucleotide adaptors contains an oligonucleotide adaptor complementary to each of the possible second overhanging ssDNA sequences of the digested cDNA.

An embodiment of Method II includes conducting further PCRs using further nesting oligonucleotide adaptors, optionally with different restriction enzymes and recognition sites, thereby sorting the amplified cDNA molecules further into non-redundant subgroups. If different restriction enzymes are used, they must cleave the ligated cDNA molecules at a point further from the ligated oligonucleotide adaptor than the overhanging ssDNA sequences of the digested cDNA, and create cohesive ends having second overhanging ssDNA sequences of a constant number of arbitrary nucleotides.

A preferred embodiment of Method II comprises repeating nesting ligation and PCR until each non-redundant subgroup contains only one type of cDNA molecule, with every expressed gene transcript in the mRNA sample uniquely represented in one of the non-redundant subgroups, i.e. "one gene per well."

Method III utilizes the non-redundant groups of cDNA fragments collected in step (5) of Method I and II for the preparation of sets of non-redundant sub-libraries of cDNA. Such sub-libraries can be more economically used for the derivation of a complete cDNA library by selecting a group of clones from the sub-libraries. The principle of Method III is that cDNA fragments derived from a specific highly abundant mRNA will converge into one group. Thus, a few groups will contain a highly redundant cDNA population. These groups are identified by analysis of the cDNA content of the group by sequencing or other methods. All other groups will be devoid of cDNAs of highly redundant mRNAs and thus of low redundancy and are used , in combination, to derive a full cDNA library. Since the elimination of the groups that contain a highly redundant cDNA population also removes some cDNA fragments of low redundancy mRNAs an approach involving parallel processing of two cDNA pools, each digested with a type different IIs restriction enzyme, is used. This makes it highly improbable that a "rare" cDNA fragment will be found in a high-redundancy group in both digest pools.

Method III broadly concerns a method of sorting genes and/or gene fragments comprising the steps of:

(1) preparing ds cDNA molecules from mRNA molecules by reverse transcription, using a poly-T primer having a general primer-template sequence upstream from the poly-T sequence that includes a recognition sequence for a restriction enzyme, yielding ds cDNA molecules having the poly-T sequence, having the general primer-template sequence;

(2) dividing the cDNA into N pools, wherein N is 1 to 25, by digesting the ds cDNA molecules with different restriction enzymes that produce digested cDNA molecules with cohesive ends having overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(3) ligating to the digested cDNA molecules of each pool a set of dsDNA oligonucleotide adaptors, each of which adaptor has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible overhanging ssDNA sequences of the digested cDNA, at the opposite end a specific primer-template sequence specific for the ssDNA adaptor

complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set;

(4) amplifying by separate PCRs the ligated cDNA molecules of each pool, utilizing for each separate PCR a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences;

(5) sorting the amplified cDNA molecules from each pool into non-redundant groups by collecting the amplification products after each separate PCR, each group of amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the PCR, wherein each of the restriction enzymes digests the N separate cDNA pools into 64 or 256 non-redundant sub-groups; and

(6) digesting cDNA fragments in each non-redundant sub-group of the cDNA pools with different restriction enzymes and further purifying the digested cDNA fragments by removing the small end fragments produced by the digestion.

In one embodiment of the method of sorting genes and/or gene fragments, the method further comprises purifying the digested cDNA fragments by removing the small end fragments produced by the digestion.

Methods I, II and III reactions stop when the cDNAs are exhausted.

In another embodiment of the method of sorting genes and/or gene fragments, the method further comprises ligating the digested cDNA fragments into a plasmid vector that has recognition sequence for a restriction enzyme and is predigested with the enzyme, producing a set of ligations.

In another embodiment of the method of sorting genes and/or gene fragments, the restriction enzyme is *Not*I or *Asc*I.

In another embodiment of the method of sorting genes and/or gene fragments, the method further comprises ligating the digested cDNA fragments into a genetic vector.

In another embodiment of the method of sorting genes and/or gene fragments, the genetic vector is a viral vector, a bacterial vector, a protozoan vector, a retrotransposon, a transposon, a DNA vector, or a recombinant vector.

In another embodiment of the method of sorting genes and/or gene fragments, the method further comprises transforming the ligation products into bacteria and growing the bacteria in a suitable growth media.

In another embodiment of the method of sorting genes and/or gene fragments, the bacteria are grown on bacteria growth plates.

In another embodiment of the method of sorting genes and/or gene fragments, N is two and the restriction enzymes of step (2) are BbsI for one pool and BsaI for the second pool.

In another embodiment of the method of sorting genes and/or gene fragments, N is 2 to 20, preferably 2 to 15, more preferably 2 to 10 and most preferably 2 to 4.

In another embodiment of the method of sorting genes and/or gene fragments, N is two and the restriction enzyme in step (1) comprises *Asc*I or another similar rare restriction enzyme.

In yet another embodiment of the method of sorting genes and/or gene fragments, N is two and the restriction enzyme in step (5) comprises *Bbs*I or *Bsa*I.

In a further embodiment of the method of sorting genes and/or gene fragments, N is two and the restriction enzyme in step (6) comprises *Not*I or *Asc*I.

This invention also provides a method of making sub-libraries of ligation sets by ligating restriction enzyme digested fragments produced by the method of sorting genes and/or gene fragments, into a plasmid vector that have recognition sequence for said restriction enzymes and predigesting with these enzymes to make 64xN or 256xN sets of ligations, wherein N is 1 to 25.

This invention further provides a method of making sub-libraries of expression system colonies by transforming the set of ligations into an expression system to produce colonies of the expression system containing each of the 64 x N or 256 x N non-redundant subgroups of cDNA fragments, wherein N is 1 to 25.

In one embodiment of the method of making sub-libraries of expression system colonies, the expression system is a bacterium.

In another embodiment of the method of making sub-libraries of expression system colonies, the bacteria are grown under suitable conditions.

In a further embodiment of the method of making sub-libraries of expression system colonies, the bacteria are plated onto bacterial growth plates.

The practice of the present invention employs, unless indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are within the skill of the art. Such techniques are explained fully in the literatures, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook et al. 1989); "Oligonucleotide Synthesis" (Gait, ed. 1984); "Animal Cell Culture" (Freshney ed., 1987) Met. Enzymol. (Academic Press, Inc.); "Handbook of Experimental Immunology" (Wei and Blackwell, eds.); "Gene Transfer Vectors for Mammalian Cells" (Miller and Calos eds. 1987); "Current Protocols in Molecular Biology" (Ausubel et al. eds. 1987); "PCR: The Polymerase Chain Reaction" (Mullis et al. eds. 1994); and "Current Protocols in Immunology" (Coligan et al. eds. 1991). These techniques are applicable to the production of the polynucleotides of the invention, and, as such, may be considered in making and practicing the invention. This invention can be applicable to the uses disclosed in PCT publications, such as WO 98/51789A2, WO 93/18176A1 and WO 99/60164.

Reference is made to U.S. Patent Nos.: 5,407,813; 5,413,909; 5,487,985; 5,508,169; 5,556,773; 5,580,726; 5,629,179; 5,650,274; 5,695,937; 5,700,644; 5,710,000; 5,728,524; 5,763,239; 5,804,382; 5,814,445; 5,837,468; 5,858,656; 5,863,722; 5,866,330; and 5,871,697; PCT publication WO 94/01582; Guilfoyle et al. (1997) Nucl. Acids Res. 25:1854-1858; Ivanova and Belyavsky (1995) Nucl. Acids Res. 23:2954-2958; Mahadeva et al. (1998) J. Mol. Biol. 284:1391-1398; Troutt et al. (1992) Proc. Natl. Acad. Sci. USA 89:9823-9825; Kato (1995) Nucl. Acids Res. 23:3685-3690; Prashar and Weissman (1996) Proc. Natl. Acad. Sci. USA 93:659-663; Ko Nucl. Acids Res. 18:5705-5711; Edward Nucl. Acids Res. 19:5227-5232; Hoog Nucl. Acids Res. 19:6123-6127; Sokolov et al. (1994) Nucl. Acids Res. 22:4009-4015; Schmidt and Mueller Nucl. Acids Res. 24:1789-1791; Belyavsky et al. (1989) Nucl. Acids Res. 17:2919-2932; Calvet (1991) Ped. Nephrol. 5:751-757; Cooke et al. (1996) Plant J. 9:101-124; Domec et al. (1990) Anal. Biochem.

188:422-426; Haymerle et al. (1986) Nucl. Acids Res. 14:8615-8625; Kato et al. (1994) Gene 150:243-250; Kohchi et al. (1995) Plant J. 8:771-776; Patanjali et al. (1991) Proc. Natl. Acad. Sci. USA 88:1943-1947; Podhajska et al. (1992) Met. Enzymol. 216:303-309; and Szybalski et al. (1991) Gene 100:13-26; and the documents cited therein and the documents of record in the prosecution of cited U.S. patent; all of which are incorporated herein by reference.

With respect to cDNAs for expression in a vector and documents providing such exogenous DNA, as well as with respect to the expression of transcription and/or translation factors for enhancing expression of nucleic acid molecule, reference is made to U.S. Patent No. 5,990,091, and WO 98/00166 and WO 99/60164, and the documents cited therein and the documents of record in the prosecution of that patent and those PCT applications; all of which are incorporated herein by reference. Thus, U.S. Patent No. 5,990,091 and WO 98/00166 and WO 99/60164 and documents cited therein and documents or record in the prosecution of that patent and those PCT applications, and other documents cited herein or otherwise incorporated herein by reference, can be consulted in the practice of this invention; and, all exogenous nucleic acid molecules and vectors cited therein can be used in the practice of this invention. In this regard, mention is also made of U.S. Patents Nos.: 6,004,777; 5,997,878; 5,989,561; 5,976,552; 5,972,597; 5,858,368; 5,863,542; 5,833,975; 5,863,542; 5,843,456; 5,766,598; 5,766,597; 5,762,939; 5,756,102; 5,756,101; and 5,494,807. The expression systems are disclosed in U.S. Patent Nos.: 5,538,885; 5,641,663; 5,830,692; and 6,004,941.

As used herein, ;vectors include, but are not limited to, viral; bacterial; protozoan; DNA; retrotransposon; transposon; or a recombinant vector thereof.

As used herein, "rare restriction enzymes" means restriction enzymes having a low chance of cleaving within the cDNA. Generally, enzymes that have recognition sequence of 8 or more base pairs are regarded as rare enzymes since they cleave, statistically, once every 4^8 bp (~1 per 16,000 bp).

EXPERIMENTAL DETAILS

The following examples illustrate some embodiments of the present invention in more detail. However, the following examples should not be construed as limiting the scope of the present invention.

Example 1

cDNA Preparation

Conversion of mRNA into ds cDNA

For priming the synthesis of single stranded cDNA from polyA+ mRNA an oligo(dT) primer was used. The primer was of the following structure, including a general primer-template sequence:

5'-TGCATGGCACAGTACTGAGTGGTATCGACTCGTACAGGCGC

GCCTTTTTTTTTTTTTTTTTV-3' (SEQ ID NO: 1) (V=C, G or A).

General primers for amplification from this sequence include GPI1, GPI2 and GPI3, SEQ ID NOs: 2, 3 and 4, respectively.

TGCATGGGA C A G T A C T G A G T

C A C A G T A C T G A G T G G T A T C G

A G T G G T A T C GACTCGTACAG

As depicted here, the three general primers are nested relative to each other.

Conventional methods were used for preparation of double stranded cDNA from polyA+ mRNA. The double stranded cDNA was column purified (Qiagen - QIAquick PCR purification kit, catalogue no. 28106) to remove excess oligo(dT) primer and nucleotides.

ds cDNA digestion; restriction enzyme choice

The double stranded cDNA was digested with a type IIs restriction enzyme (RE) that produced a four base overhang structure and that cut at least 8 nucleotides away from the recognition sequence. Other enzymes, including type II restriction enzymes, that produce other overhangs or that cut closer to the recognition sequence can be used. REs used were:

*Bbv*I 5'-GCAGCNNNNNNNN-3' (SEQ ID NO: 5)

3'-CGTCGNNNNNNNNNNNN-5' (SEQ ID NO: 6)

*Fok*I 5'-GGATGNNNNNNNNN-3' (SEQ ID NO: 33)

3'-CCTACNNNNNNNNNNNNN-5' (SEQ ID NO: 7)

In the examples as shown in Figure 4, double stranded cDNA derived from rat liver mRNA was digested with BbvI and ligated to Tail adaptor set 2. Helper oligonucleotide was HOLL2 for "Figure 4A" and HOLL1 for "Figure 4B". Ligation specificity was tested on the albumin gene, which constitutes the most abundant mRNA in rat liver. Ligation specificity was tested with a Tail primer and albumin reverse primer. The specific Tail adaptor that should ligate to the albumin gene is

Ada-ACT. All nine Tail adaptors that have one base mismatch with the albumin specific Tail adaptor were examined. Both "Figure 4A" and "Figure 4B" show separation of DNA on 1.5% agarose gel using and ethidium bromide staining was used to visualize the DNA.

In Figure 4A, helper oligonucleotide HOLL2, a perfect match to the Tail adaptor, was used. Oligonucleotide concentration was 5pmol/25µl. The correct 200bp band is observed in the specific ACT Tail primer. However, Tails ATT, AGT, ATT and TCT give a strong 200bp-albumin band. A very weak 200bp band is observed in Tail ACG, ACC and ACA. Thus, the ligation conditions used here allow frequent mis-ligations.

In Figure 4B, helper oligonucleotide HOLL1, that has a mismatch to the first nucleotide of the constant region of the Tail adaptor, was used. Oligonucleotide concentration was 2.5pmol/25µl. The 200bp albumin specific band is observed only in the Tail-ACT amplification. None of the other Tails gave the albumin band. The 500bp band observed in the ACG lane (also seen in "A") is caused by Tail-Tail amplification of an undetermined gene. Thus, the ligation conditions used here give highly specific ligation and do not allow mis-ligations.

Example 2
Differential Ligation

Adaptor design and sequence

The digested ds cDNA was ligated to a set of oligonucleotide adaptors. Two sets of adaptors were used: a set of 64 adaptors covering all 64 combinations of three of the four nucleotides of the overhang; and a set of 256 adaptors covering all 256 combinations of the four nucleotides of the overhang.

Each adaptor comprises two DNA strands: a "long" 49-51 bp strand that contain the sequence that fits into the overhang produced by the type IIs RE's; and a "short" 18-mer strand that complements the long strand up to the overhang. Three structural versions of the short strand were examined:

5'-**XYZN**GCAGGTACGTCGTACCGCGGCCGCGTGAGCTTGAGTC GCGTGGAT-3' long strand (SEQ ID NO: 8)

3'-CGTCCATGCAGCATGGCG-5' short strand 1 (SS1) (SEQ ID NO: 9)

3'-AGTCCATGCAGCATGGCG-5' short strand 2 (SS2) (SEQ ID NO: 10)

3'- CATCCATGCACCATGGCG-5' short strand 3 (SS3) (SEQ ID NO: 11)

Note that SS2 has a mismatch to the 5th nucleotide of the long strand (just after the N) and that SS3 has a mismatch to the 6th nucleotide.

The general structure of the long strand of the adaptors is as follows for the set of 64:

5'P-XYZN --- constant --- | --- Specific --- |

where each of X, Y and Z can be any of the nucleotides but are specific for each adaptor; N is a mix of all 4 nucleotides; P is a 5' phosphate; the constant region is a sequence which is common to all 64 adaptors while the specific region is specific to each of the 64 adaptors; each adaptor has a different specific sequence.

The general structure of the adaptors is as follows for the set of 256:

5'P-WXYZ --- constant --- | --- Specific --- |

where each of W, X, Y and Z can be any of the nucleotides but are specific for each adaptor; P is a 5' phosphate; the constant region is a sequence which is common to all 256 adaptors while the specific region is specific to each of the 256 adaptors, each adaptor has a different specific sequence. For each adaptor from the set of 64 and set of 256 a specific primer, complementary to the specific region of the adaptor, has been synthesized. Figure 18 shows tail adaptors set 256, which can be represented by such a general formula.

The sequences of the entire sets of 64 and 256 adaptors can be generated from the general structures for the set of 64 and the set of 256, respectively. The list of specific primers sets are shown in Figures 10 to 13.

Figure 10 shows the general structure of the primer and the primer set J2 having the general structure of the primers

5' XYZN GCAGGT ACGTCGTACC GCGGCCGC-x-x-x-x-x-x-x-x-x-x-x-x-x-3' (SEQ ID NO: 12)

Bases 4 BspMI(6) constant(10) **NotI(8)** **Tail (20)**

wherein X, Y and Z can be any of A, T, C or G.

Figure 11 shows tail primers set J2 represented by the general formula:

Tail-XYZ 5' TCCACGCGACTCAAGCTCAC (SEQ ID NO: 13)

wherein X, Y and Z in the primer name can be any of A, T, C or G. The primer sequence is different for each of the 64 different tail primers and each one of them is

a complete reverse complement to the specific region of the tail adaptor that has the same X, Y, Z.

Figure 12 shows tail primers (set number 2) represented by the general formula: TnewXYZ 5' AACGACGCGTCGCGGTACCAG (SEQ ID NO: 14) wherein X, Y and Z can be any of A, T, C or G.

Figure 13 shows tail primers set 256 represented by the general formula: TailWXYZ 5'AACGCAGTGTTCGTTCGACGA (SEQ ID NO: 15) wherein each of W, X, Y and Z can be any of A, T, C or G.

Ligation procedure

For ligations, all 64 or 256 adaptors are mixed in equal molar concentrations. Initially ligation conditions followed conventional methods. This included the use of T4 DNA ligase at 16°C and using the SS1 strand. These ligation conditions proved inadequate since ligation specificity was low; with adaptors ligating to unmatched overhangs (Figure 4a).

The following conditions gave very high ligation specificity. 100ng digested ds cDNA was placed in ligation buffer (50mM Tris-HCl, pH 7.8; 10mM $MgCl_2$; 10mM dithiothreitol, 26μM NAD+; 25μg/ml bovine serum albumin). Adaptor concentration was 2.5pmol/12μl (long strand at 2.5pmol/12μl and the short strand at 10pmol/12μl). Importantly, short strand SS2 with one mismatch to the 5th nucleotide of the long strand (just after the N) was used. Other short strands always gave lower specificity. At this point reaction volume was 10μl. The reaction was heated to 65°C for 5 minutes and then cooled to 8°C. 2μl of *E. coli* DNA ligase (10units/μl) were added.

Incubation was carried out for 12 hours. The reaction was stopped by heating to 65°C for 15 minutes and the reaction mix was stored at 4°C. Ligation products were column purified (QIAquick spin) to remove unligated adaptors.

Example 3

Analysis of ligation specificity

Ligation specificity was tested on highly expressed genes. The following example details an experiment performed on mRNA from rat liver. The most abundant gene in this tissue is albumin and was selected (as well as other genes not shown here) to test ligation specificity. The type IIs RE used was *Bbv*I. The 3'-most

*Bbv*I site in the rat albumin gene (GenBank accession no. J00698) is at nucleotide 1740, 250bp from the poly-A tail.

A reverse oligonucleotide 5'-CACCAACAGAAGAGATGAGTCCTG-3' (SEQ ID NO: 16) matches nucleotides 1901 to 1881. The distance of this oligonucleotide from the *Bbv*I site is 160bp.

The specific adaptor for ligation to this *Bbv*I end of the rat albumin gene is Ada-ACT: 3'AGATGCGGATCGGGCTCTGTGCGCCGGCGCCATGCTG CATGGACGNTCA5' (SEQ ID NO:17)

Amplification of the ligation product with specific-ACT (5'-TCTACGCCTAGCCCGAGACAC-3' (SEQ ID NO: 18)) by PCR gave the correct fragment size of 209bp on an agarose gel (Figure 4 lane ACT). Had a different adaptor managed to ligate (i.e. mis-ligate) to the end of albumin cDNA, then a different specific primer would have given a fragment of the same size. Figure 4a shows the results of ligation done under non-specific conditions using a short strand with no mismatches. Lanes ATT, AGT, AAT, TCT show the presence of such a fragment after amplification with other tail primers indicating presence of mis-ligation. However, when the conditions defined above were used, no mis-ligations occurred (Figure 4B).

Additional experiments performed on the GAPDH sequence included testing ligation specificity on all 64 specific adaptors. Upon digestion with a type IIs restriction enzyme of a double-stranded cDNA derived from the mRNA of a specific gene, fragments with specific overhangs are produced. The example below describes the full human GAPDH cDNA sequence and the location of the recognition sites for the BbvI type IIs restriction enzyme. The "><" symbol marks the exact point were the enzyme cleaves the cDNA. The polyA addition signal (AATAAA), found 20 to 30 bases before the actual polyA addition site, is underlined. Also underlined, in the more upstream regions, are the BbvI recognition sequences. The example given here is in addition to the rat albumin example.

GTTCGACAGTCAGCCGCATCTTCTTTTGCGTCGCCAGCCGAGCCACATCG
CTCAGACACCATGGGGAAGGTGAAGGTCGGAGTCAACGGATTTGGTCGT
ATTGGGCGCCTGGTCACCAGGGCTGCTTTTAACTCTGGTAAAGTGGATAT
TGTTGCCATCAATGACCCCTTCATTGACCTCAACTACATGGTTTACATGT

TCCAATATGATTCCACCCATGGCAAATTCCATGGCACCGTCAAGGCTGA
GAACGGGAAGCTTGTCATCAATGGAAATCCCATCACCATCTTCCAGGAG
CGAGATCCCTCCAAAATCAAGTGGGGCGATGCTGGCGCTGAGTACGTCG
TGGAGTCCACTGGCGTCTTCACCACCATGGAGAAGGCTGGGGCTCATTT
GCAGGGGGGAGCCAAAAGGGTCATCATCTCTGCCCCCTCTGCTGATGCC
CCCATGTTCGTCATGGGTGTGAACCATGAGAAGTATGACAACAGCCTCA
AGATCATCAGCAATGCCTCCTGCACCACCAACTGCTTAGCACCCCTGGCC
AAGGTCATCCATGACAACTTTGGTATCGTGGAAGGACTCATGACCACAG
TCCATGCCATCACTGCCACCCAGAAGACTGTGGATGGCCCCTCCGGGAA
ACTGTGGCGTGATGGCCGCGGGGCTCTCCAGAACATCATCCCTGCCTCTA
CTGGCGCTGCCAAGGCTGTGGGCAAGGTCATCCCTGAGCTGAACGGGAA
GCTCACTGGCATGGCCTTCCGTGTCCCCACTGCCAACGTGTCAGTGGTGG
ACCTGACCTGCCGTCTAGAAAAACCTGCCAAATATGATGACATCAAGAA
GGTGGTGAAGCAGGCGTCGGAGGGCCCCCTCAAGGGCATCCTGGGCTAC
ACTGAGCACCAGGTGGTCTCCTCTGACTTCAACAGCGACACCCACTCCTC
CACCTTTGACGCTGGGGCTGGCATTGCCCTCAACGACCACTTTGTCAAGC
TCATTTCCTGGTATGACAACGAATTTGGCTACAGCAACAGGGTGGTGGA
CCTCATGGCCCACATGGCCTCCAAGGAGTAAGACCCCTGGACCACCAGC
CCCAGCAAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGT
CCCTGCCACACTCAGTCCCCCACCACACTGAATCTCCCCTCCTCACAGTT
GCCATGTAGACCCCTTGAAGAGGGGAGGGGCCTAGGGAGCCGCACCTTG
TCATGTACCATCAATAAAGTACCCTGTGCTCAACC (SEQ ID NO: 19)

The expected end of the 3' fragment is:

5'-AAGTGTTGCAAGGCTGCCGACAAGGATAAC-3' (SEQ ID NO: 20)

3'-CAACGTTCCGACGGCTGTTCCTATTG-5' (SEQ ID NO: 34)

The cDNA derived in the SDGI procedure has an extended polyA tail of a specific sequence. This is underlined in the sequence below which describes the exact structure of the double stranded structure of the 3' most fragment of the human GAPDH cDNA. Note the overhang structure of the 5' end. The *Bbv*I recognition sequence is underlined.

GCCTCTACTGGCGCTGGATGACCGCGACCCAAGGCTGTGGGCAAGGTCA
TCCCTGAGCTGAACGGGAAGCTCACTGGCATGGCCTTCCGTGTCCCCACG

GTTCCGACACCCGTTCCAGTAGGGACTCGACTTGCCCTTCGAGTGACCGT
ACCGGAAGGCACAGGGGTGTGCCAACGTGTCAGTGGTGGACCTGACCTG
CCGTCTAGAAAAACCTGCCAAATATGATGACATCAAGAAGACGGTTGCA
CAGTCACCACCTGGACTGGACGGCAGATCTTTTTGGACGGTTTATACTAC
TGTAGTTCTTCGTGGTGAAGCAGGCGTCGGAGGGCCCCCTCAAGGGCAT
CCTGGGCTACACTGAGCACCAGGTGGTCTCCTCACCACTTCGTCCGCAGC
CTCCCGGGGGAGTTCCCGTAGGACCCGATGTGACTCGTGGTCCACCAGA
GGACTGACTTCAACAGCGACACCCACTCCTCCACCTTTGACGCTGGGGCT
GGCATTGCCCTCAACGACCACTTGACTGAAGTTGTCGCTGTGGGTGAGG
AGGTGGAAACTGCGACCCCGACCGTAACGGGAGTTGCTGGTGAATGTCA
AGCTCATTTCCTGGTATGACAACGAATTTGGCTACAGCAACAGGGTGGT
GGACCTCATGGCCCACACAGTTCGAGTAAAGGACCATACTGTTGCTTAA
ACCGATGTCGTTGTCCCACCACCTGGAGTACCGGGTGATGGCCTCCAAG
GAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGAAGA
GAGAGACCCTTACCGGAGGTTCCTCATTCTGGGGACCTGGTGGTCGGGG
TCGTTCTCGTGTTCTCCTTCTCTCTCTGGGACACTGCTGGGGAGTCCCTGC
CACACTCAGTCCCCCACCACACTGAATCTCCCCTCCTCACAGTTGCCATG
GTGACGACCCCTCAGGGACGGTGTGAGTCAGGGGGTGGTGTGACTTAGA
GGGGAGGAGTGTCAACGGTACTAGACCCCTTGAAGAGGGGAGGGGCCT
AGGGAGCCGCACCTTGTCATGTACCATCAATAAAGTACCCTGTATCTGG
GGAACTTCTCCCCTCCCCGGATCCCTCGGCGTGGAACAGTACATGGTAGT
TATTTCATGGGACAGCTCAACCAAAAAAAAAAAAAAAAAACGAGTTCCT
TTTTTTTTTTTTTTTT (SEQ ID NO: 21)

The specific adaptor that will ligate to this overhang is:

GGTACGACGTTCAGCAGCCTCTACTGGCGCTG (SEQ ID NO: 35)

Ada AGG 3' CCAATAGGCAGCCGCCGCTGCCATGCTGCAAGTCGANGGAGA
ACCGCGAC (SEQ ID NO: 22)

Note the mismatch in the upper sequence (helper) of the adaptor, marked by an underline. To the right of the adaptor, the end of the human GAPDH sequence is shown to emphasize the match between the adaptor and the overhang. Ligation specificity is examined by the ability of the "TAIL" primer that matches the 3' (specific) part of the adaptor (Tail AGG 5' GGTTATCCGTCGGCGGCGAC 3')

(SEQ ID NO: 23) to amplify, in combination with a GAPDH-specific reverse primer (underlined above – 5' TACAGCAACAGGGTGGTGGA 3') (SEQ ID NO:24). This PCR amplification should result in a fragment of a specific size, 390bp in the example of GAPDH (350 + 40 of the adaptor). Complete specificity is achieved when all of the other TAIL primers are unable to amplify the GAPDH sequence. This is what is shown in Figure 5 where all 64 TAIL plus GAPDH reverse PCR amplifications were performed and only the Ada-AGG TAIL gave the expected fragment of 390bp.

Example 4
Ligation Efficiency Analysis

To examine ligation efficiency, the successful amplification of a set of rare mRNAs was tested. As above, reverse primers for the specific genes are used in combination with the specific primers that were expected to ligate to the ends of these cDNAs. All reaction conditions were performed as described above.

Example 5
Amplification-division of the Different Groups (general-specific PCR)

The ligation could employ a mix of all 64 (or 256) adaptors. While the following details the protocol performed on the set of 64 adaptors, the same protocol applies to a set of 256 adaptors. To divide the ligated cDNA into 64 groups, 64 PCR reactions were performed. Each reaction used a primer specific for one of the specific ligation adaptors, and the general primer. This resulted in a specific amplification of all cDNAs ligated with the specific adaptor.

PCR conditions were: 2 min. at 95°C; followed by 30 cycles of 1 min. at 95°C; 1 min. at 58°C; and 2 min. at 68°C, followed by incubation for 7 min. at 68°C. Figure 6 shows the products of the 64 specific-general-primed PCR reaction. Southern blot analysis of the 64 reactions (Figure 7) demonstrates the specificity of the procedure. After amplification with specific and general primers, GAPDH mRNA was amplified only in the expected group (AGG).

The PCR products were column purified (QIAquick spin) to remove the unincorporated primers and nucleotides. In this step, an mRNA source of 10,000 genes was divided into 64 groups each containing an average of 150 cDNA species (genes). If the source contains all 100,000 human genes, each group will contain an average of 1500 cDNA species (genes).

Example 6
Priming Nesting Procedure

First nesting

In this step, each of the 64 groups from the previous step was further sorted into 256 subgroups. Division into fewer groups is also possible. A set of nesting oligonucleotides, the 1st nesting set, was used. This set of 256 nesting primers could be used on all 64 general-specific primed groups (as well as on the 256 general-specific primed groups) since they prime from the "constant" region of the specific adaptor. The overall structure of a 1st nesting primer is: 5'-GCGGCCGCGGTACGACGTACCTGCNIIIWXYZ-3' (SEQ ID NO: 25) where I=inosine; N=any nucleotide; each of W, X, Y, Z=C, G, T or A.

The NIII nucleotides match the four nucleotides in the specific adaptor used to ligate to the overhang end of the cDNAs. The inosine nucleotides can match any of the regular nucleotides. The WXYZ nucleotides, covering all 256 possibilities of C, G, T or A allow nesting into the four nucleotides adjacent to the overhang. The first nesting oligonucleotide list is shown in Figures 14 and 15.

Figure 14 shows first nesting primers 256 for tail adaptor 64 set 1, represented by the formula: 5' GGTACGACGTTCAGCTNIIIWXYZ (SEQ ID NO: 26) wherein W, X, Y and Z can be any of A, T, C or G.

Figure 15 shows first nesting primers 64 for tail adaptor 64 set 1, represented by the formula: 5' GGTACGACGTTCAGCTNIIIXYZ (SEQ ID NO: 27) wherein X, Y and Z can be any of A, T, C or G.

An optional λ exonuclease reaction can be performed to eliminate carry-over of cDNA from the original cDNA reaction. This is because the oligo(dT) primer used to produce the cDNA in the reverse transcription reaction is phosphorylated and the general and specific primers used for general-specific primed amplifications are not phosphorylated. The following mixture was prepared: 2μl of purified general-specific primed PCR product; 6μl H_20, 1μl λ exonuclease buffer; and 1μl λ exonuclease. The reaction mixture was then column purified.

For nesting, a 1:500 dilution of the general-specific PCR product was taken. PCR reaction constituents were standard (including anti-Taq antibody). Cycling conditions were: 1 min. at 95°C; 1 min. at 59°C; and 2 min. at 70°C. 30 cycles

were performed. After PCR, the unincorporated primers and nucleotides were removed using QIAquick spin columns.

The 1st nesting stage divides each of the 64 groups into 256 groups for a total of 16,384 groups. Thus, for an mRNA source of 10,000 genes, each of the 256 1st nesting tubes should contain an average of less than 1 cDNA species (gene). This means that most tubes (>100) will contain one cDNA species, some will be empty and a few will contain more then one cDNA species. Figure 8 shows the results of a 1st nesting PCR done on 3 of the 64 groups. The object of the nesting PCR was to isolate three specific genes according to the sequences around the *Bbv*I site closest to the 3' end.

For a source containing all 100,000 human genes, each of the 256 tubes will contain an average of 6 cDNA species (genes). Thus, a further nesting round would achieve one gene only per well.

Second nesting

In this stage, each of the 256 1st nesting groups was further divided into 16 groups. As for the 1st nesting primers, this set of 16 2nd nesting primers can be used on all 1st nesting primer reactions, since they prime from the "constant" region of the ligation adaptors.

The primers used for the 2nd nesting are of the structure:
5'GCGGCCGCGGTACGACGTACCTGCNGGGIIIINNXY3' (SEQ ID NO: 28)
where I=inosine; N=any nucleotide; each of W, X, Y and Z can be any of C, G, T or A. In places were inosine was present in the 1st nesting primer a "G" is placed in the 2nd nesting primers (since a "C" is incorporated as a match to "I"). Lists detailing second nesting primers are shown in Figures 16 and 17.

Figure 16 shows second nesting primers 64 for tail adaptor 64 set 1, represented by the general formula: 5' GGTACGACGTTCAGCTNGGG IIIXYZ (SEQ ID NO: 29) wherein each of X, Y and Z can be any of A, T, C or G.

Figure 17 shows second nesting primers 16 for tail adaptor 64 set 1, represented by the general formula: 5' GGTACGACGTTCAGCTNGGG IIIXY (SEQ ID NO: 30) wherein each of X and Y can be any of A, T, C or G.

A 1:500 proportion of the 1st nesting purified PCR products are used for a PCR reaction performed with exactly the same constituents and conditions as

described above. Figure 9 shows the results of 2nd nesting PCR on the three groups shown in Figure 8. Highly pure DNA fragments were obtained.

The 2nd nesting stage divides each of the 16,384 1st nesting groups into sixteenths, for a total of 262,144 groups. Thus, about 100,000 of the groups should contain cDNA products and more than 95% of them should contain only one gene or gene fragment.

Example 7
Ligation Nesting Procedure

Digestion of END-TAIL PCR product

The column purified PCR products of each of the 64 general-specific primed groups described above were digested with *Bsp*MI under standard manufacturer (New England BioLabs) conditions. The released adaptors were removed by column purification (QIAquick spin).

First nesting ligation (Adaptor set #2)

100ng of each of the 64 digested products were ligated with a mix of 64 adaptors (nesting ligation adaptor set). Ligation conditions were identical to those detailed above with differential ligation. The same specificity and efficiency tests, detailed above, were successfully performed. Each ligation was column purified (QIAquick spin) to remove unligated adaptors. DNA was eluted from the column in a final volume of 100µl. Adaptor set #2 is shown in Figure 20. The tail adaptors 64 (set number 2) in Figure 20 can be represented by the general formula:

|specific|----------constant|----------specific--------|

5' Ph-XYZN GCAGGTACGTCGTACC **GCGGCCGC**

GTGAGCTTGAGTCGCGTGGA (SEQ ID NO: 31) wherein X, Y and Z can be any of A, T, C or G.

Amplification of the first ligation products

Each of the 64 ligations was then divided into 64 tubes. The final number of tubes was thus 4096. From each ligation tube 1µl was taken for each of the 64 amplifications. Each amplification was done by one of the 64 specific primers and the general primer. Amplification conditions were identical to those detailed in the "Amplification-division of the different groups (general-specific primed PCR)" described above. Each PCR reaction mixture was column purified (QIAquick spin) to remove unincorporated primers and nucleotides.

Digestion of END-TAII first ligation PCR product

The procedure detailed in the "Digestion of general-specific primed PCR product" above is repeated.

Second nesting ligation

The procedure detailed for the 1st nesting ligation is repeated with adaptor set #1. Figure 19. The tail adaptors 64 (set number 1) of Figure 19 are represented by the formula:

|------constant-----|------specific---------|

AdaXYZ 5'PhTTTNAGCTGAACGTCGTACCCGTCGAACGAACACGGGCGT (SEQ ID NO: 32) wherein each of X, Y and Z can be any of A, T, C or G.

Example 8

Gene Analysis (Agarose Gel and Sequencing)

PCR products obtained from the 2nd nesting reaction (either priming or ligation) are separated an agarose gel to examine the presence of PCR products and the number of fragments (Figure 9). Sorted or isolated cDNAs are purified and sequenced using the constant region of the ligation adaptors as a primer.

Example 9

Construction of cDNA Library from the Amplification Products Obtained from the Differential Ligation Step

The double stranded cDNA, prepared as described above, was divided into two pools. One pool was digested with *Bbs*I and the second with *Bsa*I. The following procedure was done in parallel for each pool.

Following differential ligation, performed as described above, using adaptor set J2 for ligation, PCR amplification was performed as with primer set J2. Amplified products were column purified (QIAquick spin). The PCR products from each of the 64 groups were digested with *Not*I and *Asc*I and were column purified (QIAquick spin). A plasmid that contains *Not*I and *Asc*I in its multiple cloning site is digested with *Not*I and *Asc*I and the linearized fragment is purified. The purified *Not*I-*Asc*I digested products are then ligated to a linearized plasmid.

Ligation products were transformed into bacteria using standard protocols. Transformed bacteria were plated onto growth plates and, following standard incubation, hundreds to thousands of colonies grow on each plate. For sequencing, each plasmid was purified from picked colonies and prepared for sequencing using

standard protocols. Another option is to amplify the insert from the plasmid found in the picked colonies using primers flanking the insert. The amplified inserts are sequenced using standard protocols.

The double stranded cDNA, prepared as described above, is divided into between 3 and 25 pools for digestion with restriction enzymes, ligation and expression. Expression of the separated genes can be in a bacterium.

Having thus described in detail preferred embodiments of the present invention, it is to be understood that the invention defined by the appended claims is not to be limited by particular details set forth in the above description as many apparent variations thereof are possible without departing from the spirit or scope thereof.

REFERENCES

Belyavsky et al. (1989) Nucl. Acids Res. 17:2919-2932
Calvet (1991) Ped. Nephrol. 5:751-757
Cooke et al. (1996) Plant J. 9:101-124
Domec et al. (1990) Anal. Biochem. 188:422-426
Edward Nucl. Acids Res. 19:5227-5232
Guilfoyle et al. (1997) Nucl. Acids Res. 25:1854-1858
Haymerle et al. (1986) Nucl. Acids Res. 14:8615-8625
Hoog Nucl. Acids Res. 19:6123-6127
Ivanova and Belyavsky (1995) Nucl. Acids Res. 23:2954-2958
Kato et al. (1994) Gene 150:243-250
Kato (1995) Nucl. Acids Res. 23:3685-3690
Ko Nucl. Acids Res. 18:5705-5711
Kohchi et al. (1995) Plant J. 8:771-776
Mahadeva et al. (1998) J. Mol. Biol. 284:1391-1398
Patanjali et al. (1991) Proc. Natl. Acad. Sci. USA 88:1943-1947
Podhajska et al. (1992) Met. Enzymol. 216:303-309
Prashar and Weissman (1996) Proc. Natl. Acad. Sci. USA 93:659-663
Schmidt and Mueller Nucl. Acids Res. 24:1789-1791
Sokolov et al. (1994) Nucl. Acids Res. 22:4009-4015
Szybalski et al. (1991) Gene 100:13-26
Troutt et al. (1992) Proc. Natl. Acad. Sci. USA 89:9823-9825
Unrau and Deugau (1994) Gene 145:163-169

United States Patents

5,407,813	5,756,101	5,858,656
5,413,909	5,756,102	5,863,542
5,487,985	5,763,239	5,863,542
5,494,807	5,766,598	5,863,722
5,538,885	5,766,597	5,866,330
5,556,773	5,762,939	5,871,697
5,629,179	5,804,382	5,972,597
5,641,663	5,814,445	5,976,552
5,650,274	5,830,692	5,989,561
5,700,644	5,833,975	5,990,091
5,707,807	5,837,468	5,997,878
5,710,000	5,843,456	6,004,777
5,728,524	5,858,368	6,004,941

PCT publications

WO 93/18176A1	WO 94/01582	WO 98/00166	WO 98/51789	WO 99/60164

CLAIMS

1. A method of sorting genes comprising:

(1) preparing ds cDNA molecules from mRNA molecules by reverse transcription, using a poly-T primer optionally having a general primer-template sequence upstream from the poly-T sequence, yielding ds cDNA molecules having the poly-T sequence, optionally having the general primer-template sequence;

(2) digesting the ds cDNA molecules with a restriction enzyme that produces digested cDNA molecules with cohesive ends having overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(3) ligating to the digested cDNA molecules a set of dsDNA oligonucleotide adaptors, each of which adaptor has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible overhanging ssDNA sequences of the digested cDNA, at the opposite end a specific primer-template sequence specific for the ssDNA adaptor complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set;

(4) amplifying by separate PCRs the ligated cDNA molecules, utilizing for each separate PCR a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences; and

(5) sorting the amplified cDNA molecules into non-redundant groups by collecting the amplification products after each separate PCR, each group of amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the PCR.

2. The method according to claim 1, wherein the restriction enzyme is selected from type IIs restriction enzymes.

3. The method according to claim 2, wherein the type IIs restriction enzyme is *Bbv*I, *Bsp*MI, *Fok*I, *Hga*I, *Mbo*I, *Bbs*I, *Bsa*I, *Nsp*MI, *Bsm*BI or *Sfa*NI.

4. The method according to claim 1, wherein the restriction enzyme is selected from type II restriction enzymes.

5. The method according to claim 4, wherein the from type II restriction enzyme is *Bgl*I, *Bst*XI or *Sfi*I.

6. The method according to claim 1, wherein a complete set of oligonucleotide adaptors and specific primers contains an oligonucleotide adaptor and a specific primer complementary to each of the possible overhanging ssDNA sequences of the digested cDNA.

7. The method according to claim 1 wherein the 3'-most nucleotide of the ssDNA complementary sequence of the oligonucleotide adaptor is an arbitrary nucleotide N, which pairs with the 5'-most nucleotide of each of the possible overhanging ssDNA sequences of the digested cDNA.

8. The method according to claim 7, comprising using a complete set of oligonucleotide adaptors and specific primers, containing an oligonucleotide adaptor and a specific primer complementary to each of the possible overhanging ssDNA sequences of the digested cDNA excluding the 5'-most nucleotide that pairs with the arbitrary nucleotide N of the oligonucleotide adaptor.

9. The method according to claim 8, wherein a complete set of oligonucleotide adaptors have 4, 16, 64, 256, or 1024 oligonucleotide adaptors; wherein the constant number of arbitrary nucleotides is 1, 2, 3, 4, or 5.

10. The method according to claim 1 further comprising:

(1) amplifying the sorted non-redundant groups of cDNA molecules by nesting PCR, each amplification utilizing a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template sequence, as well as one of a set of nesting primers with the following general formula:

5'-|sequence complementary to the constant sequence of the oligonucleotide adaptors|-NI_x-|1-5 nucleotides complementary to one of the possible sequences of 1-5 nucleotides immediately upstream from the overhanging ssDNA sequence on the cDNA|-3' where N is an arbitrary nucleotide; I is inosine; and x=1,2,3 or 4, being one fewer

than the constant number of nucleotides in the overhanging ssDNA sequences; and

(2) sorting the amplified cDNA molecules into non-redundant subgroups by collecting the amplification products after each separate nesting PCR, each non-redundant subgroup of cDNA molecules determined by the particular nested primer that complemented the 1-5 nucleotides immediately upstream from the overhanging ssDNA sequence on the cDNA.

11. The method according to claim 10 comprising using a complete set of nesting primers, containing a nesting primer complementary to each of the possible sequences of 1-5 nucleotides immediately upstream from the overhanging ssDNA sequence on the cDNA.

12. The method according to claim 10, comprising conducting further PCRs with further nesting primers complementary to the next immediately upstream cDNA nucleotides, thereby sorting the amplified cDNA molecules further into non-redundant subgroups.

13. The method according to claim 12, further comprising repeating the steps according to claim 10 until each non-redundant subgroup contains only one type of cDNA molecule, with every expressed-gene transcript in the mRNA sample uniquely represented in one of the non-redundant subgroups.

14. A method of sorting genes comprising:

(1) preparing ds cDNA molecules from mRNA molecules by reverse transcription, using a poly-T primer optionally having a general primer-template sequence upstream from the poly-T sequence, yielding ds cDNA molecules having the poly-T sequence, optionally having the general primer-template sequence;

(2) digesting the ds cDNA molecules with a first restriction enzyme that produces digested cDNA molecules with cohesive ends having first overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(3) ligating to the digested cDNA molecules a set of dsDNA oligonucleotide adaptors, each of which adaptor has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible first overhanging ssDNA sequences of the digested cDNA, at the opposite

end a specific primer-template sequence specific for the ssDNA adaptor complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set, and that contains a recognition site for a second restriction enzyme that can cleave the ligated cDNA molecules at a point further from the ligated oligonucleotide adaptor than the overhanging ssDNA sequences of the digested cDNA, and can create cohesive ends having second overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(4) amplifying by separate PCRs the ligated cDNA molecules, utilizing for each separate PCR a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences; and

(5) sorting the amplified cDNA molecules into non-redundant groups by collecting the amplification products after each separate PCR, each group of amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the PCR.

15. The method of claim 14 wherein the first restriction enzyme is selected from type II and type IIs restriction enzymes

16. The method according to claim 14 wherein the second restriction enzyme is selected from type IIs restriction enzymes.

17. The method according to claim 14 comprising using a complete set of oligonucleotide adaptors and specific primers, containing an oligonucleotide adaptor and a specific primer complementary to each of the possible first overhanging ssDNA sequences of the digested cDNA.

18. The method according to claim 14 wherein the 3'-most nucleotide of the ssDNA complementary sequence of the oligonucleotide adaptor is an arbitrary nucleotide N, which pairs with the 5'-most nucleotide of each of the possible first overhanging ssDNA sequences of the digested cDNA.

19. The method according to claim 18 comprising using a complete set of oligonucleotide adaptors and specific primers, containing an oligonucleotide adaptor and a specific primer complementary to each of the possible first overhanging

ssDNA sequences of the digested cDNA excluding the 5'-most nucleotide that pairs with the arbitrary nucleotide N of the oligonucleotide adaptor.

20. The method according to claim 14 further comprising:

(1) digesting the sorted non-redundant groups of cDNA molecules with the second restriction enzyme, cleaving the ligated cDNA molecules at a point further from the ligated oligonucleotide adaptor than the overhanging ssDNA sequences of the digested cDNA, and creating cohesive ends having second overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(2) ligating to the digested cDNA molecules a set of nesting dsDNA oligonucleotide adaptors, each of which adaptor has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible second overhanging ssDNA sequences of the digested cDNA, at the opposite end a specific primer-template sequence unique for the ssDNA adaptor complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set, and that contains the recognition site for the second restriction enzyme;

(3) amplifying by separate PCRs the ligated cDNA molecules, utilizing for each separate PCR a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences; and

(4) sorting the amplified cDNA molecules into non-redundant subgroups by collecting the amplification products after each separate PCR, each subgroup of amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the PCR.

21. The method according to claim 20 comprising using a complete set of nesting dsDNA oligonucleotide adaptors, containing an oligonucleotide adaptor complementary to each of the possible second overhanging ssDNA sequences of the digested cDNA.

22. The method according to claim 20, further comprising conducting further PCRs using further nesting oligonucleotide adaptors, optionally with different restriction enzymes and recognition sites, thereby sorting the amplified cDNA molecules further into non-redundant subgroups.

23. The method according to claim 22, further comprising repeating the steps according to claim 20 until each non-redundant subgroup contains only one type of cDNA molecule, with every expressed gene in the mRNA sample uniquely represented in one of the non-redundant subgroups.

24. A method of sorting genes and/or gene fragments comprising the steps of:

(1) preparing ds cDNA molecules from mRNA molecules by reverse transcription, using a poly-T primer having a general primer-template sequence upstream from the poly-T sequence that includes a recognition sequence for a restriction enzyme, yielding ds cDNA molecules having the poly-T sequence, having the general primer-template sequence;

(2) dividing the cDNA into N pools, wherein N is 1 to 25, by digesting the ds cDNA molecules with different restriction enzymes that produce digested cDNA molecules with cohesive ends having overhanging ssDNA sequences of a constant number of arbitrary nucleotides;

(3) ligating to the digested cDNA molecules of each pool a set of dsDNA oligonucleotide adaptors, each of which adaptor has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible overhanging ssDNA sequences of the digested cDNA, at the opposite end a specific primer-template sequence specific for the ssDNA adaptor complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set;

(4) amplifying by separate PCRs the ligated cDNA molecules of each pool, utilizing for each separate PCR a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences;

(5) sorting the amplified cDNA molecules from each pool into non-redundant groups by collecting the amplification products after each separate PCR, each group of amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the PCR, wherein each of the restriction enzymes digests the N separate cDNA pools into 64 or 256 non-redundant sub-groups; and

(6) digesting cDNA fragments in each non-redundant sub-group of the cDNA pools with different restriction enzymes.

25. The method according to claim 24 further comprising purifying the digested cDNA fragments by removing the small end fragments produced by the digestion.

26. The method according to claim 25 further comprising ligating the digested cDNA fragments into a plasmid vector that has recognition sequence for a restriction enzyme and is predigested with the enzyme, producing a set of ligations.

27. The method according to claim 26, wherein the restriction enzyme is NotI or AscI.

28. The method according to claim 25 further comprising ligating the digested cDNA fragments into a genetic vector.

29. The method according to claim 28, wherein the genetic vector is a viral vector, a bacterial vector, a protozoan vector, a retrotransposon, a transposon, a DNA vector, or a recombinant vector.

30. The method according to 26 further comprising transforming the ligation products into bacteria and growing the bacteria under suitable conditions.

31. The method according to claim 30, wherein the bacteria are grown on bacteria growth plates.

32. The method according to claim 24, wherein N is two and the restriction enzymes of step (2) are BbsI for one pool and BsaI for the second pool.

33. The method according to claim 24, wherein N is two and the restriction enzyme in step (1) comprises AscI or another similar rare restriction enzyme.

34. The method according to claim 24, wherein N is two and the restriction enzyme in step (5) comprises *Bbs*I or *Bsa*I.

35. The method according to claim 24, wherein N is two and the restriction enzyme in step (6) comprises NotI or AscI.

36. A method of making sub-libraries of ligation sets by ligating restriction enzyme digested fragments according to claim 24 into a plasmid vector that have recognition sequence for said restriction enzymes and predigesting with these enzymes to make 64xN or 256xN sets of ligations, wherein N is 1 to 25.

37. A method of making sub-libraries of bacterial colonies, wherein the set of ligations according to claim 26 are transformed into an expression system to produce colonies of the expression system containing each of the 64xN or 256xN non-redundant subgroups of cDNA fragments, wherein N is 1 to 25.

38. The method according to claim 37, wherein the expression system is a bacterium.

39. The method according to claim 38, wherein the bacteria are placed in a suitable growth media.

40. The method according to claim 39, wherein the growth media is bacterial growth plates.

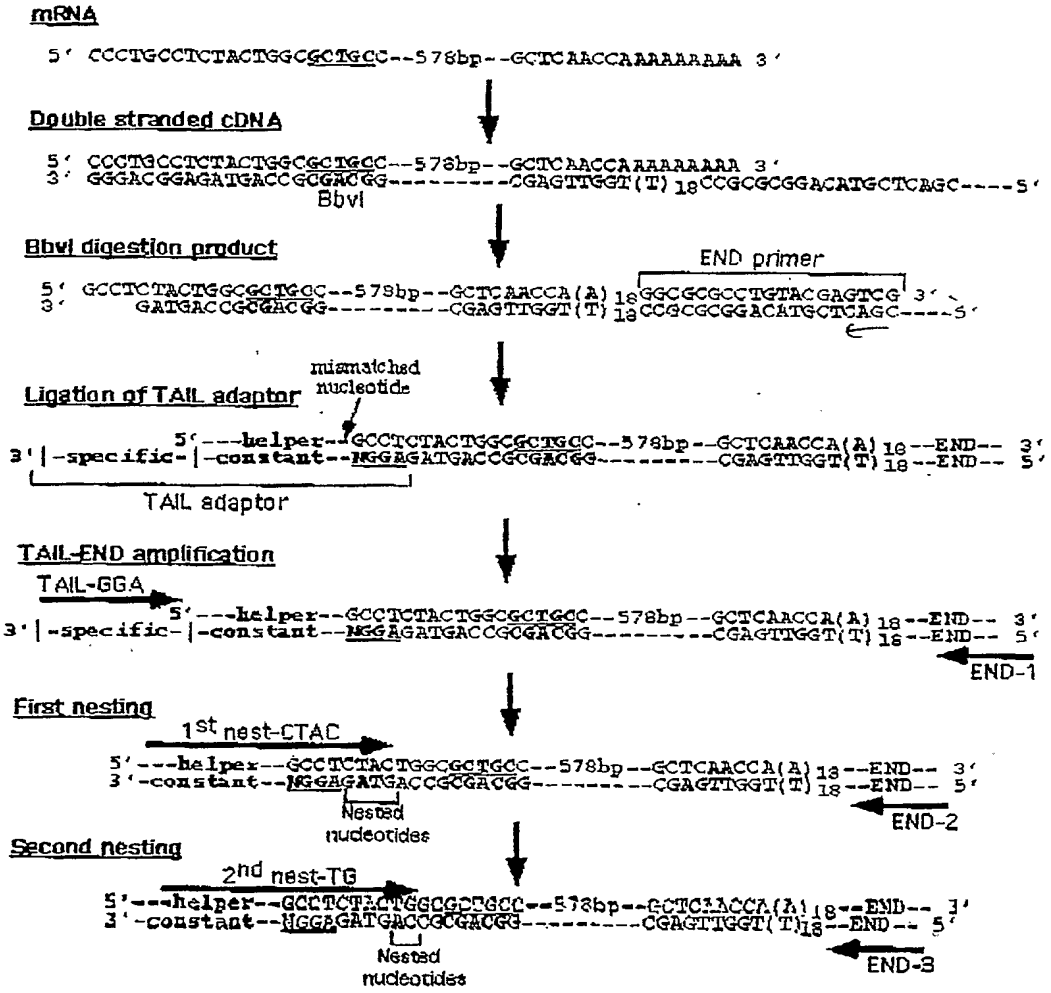

mRNA
5' CCCTGCCTCTACTGGCGCTGCC--578bp--GCTCAACCAAAAAAAAA 3'
Double stranded cDNA
5' CCCTGCCTCTACTGGCGCTGCC--578bp--GCTCAACCAAAAAAAAA 3'
3' GGGACGGAGATGACCGCGACGG---------CGAGTTGGT(T)18CCGCGCGGACATGCTCAGC----5'
BbvI
Bbvi digestion product
END primer
5' GCCTCTACTGGCGCTGCC--578bp--GCTCAACCA(A)18GGCGCGCCTGTACGAGTCG 3'
3' GATGACCGCGACGG---------CGAGTTGGT(T)18CCGCGCGGACATGCTCAGC----5'
mismatched nucleotide
Ligation of TAIL adaptor
5'---helper--GCCTCTACTGGCGCTGCC--578bp--GCTCAACCA(A)18--END-- 3'
3'|-specific-|-constant--NGGAGATGACCGCGACGG---------CGAGTTGGT(T)18--END-- 5'
TAIL adaptor
TAIL-END amplification
TAIL-GGA
5'---helper--GCCTCTACTGGCGCTGCC--578bp--GCTCAACCA(A)18--END-- 3'
3'|-specific-|-constant--NGGAGATGACCGCGACGG---------CGAGTTGGT(T)18--END-- 5'
END-1
First nesting
1st nest-CTAC
5'---helper--GCCTCTACTGGCGCTGCC--578bp--GCTCAACCA(A)18--END-- 3'
3'-constant--NGGAGATGACCGCGACGG---------CGAGTTGGT(T)18--END-- 5'
Nested nucleotides
END-2
Second nesting
2nd nest-TG
5'---helper--GCCTCTACTGGCGCTGCC--578bp--GCTCAACCA(A)18--END-- 3'
3'-constant--NGGAGATGACCGCGACGG---------CGAGTTGGT(T)18--END-- 5'
Nested nucleotides
END-3

FIG. 1

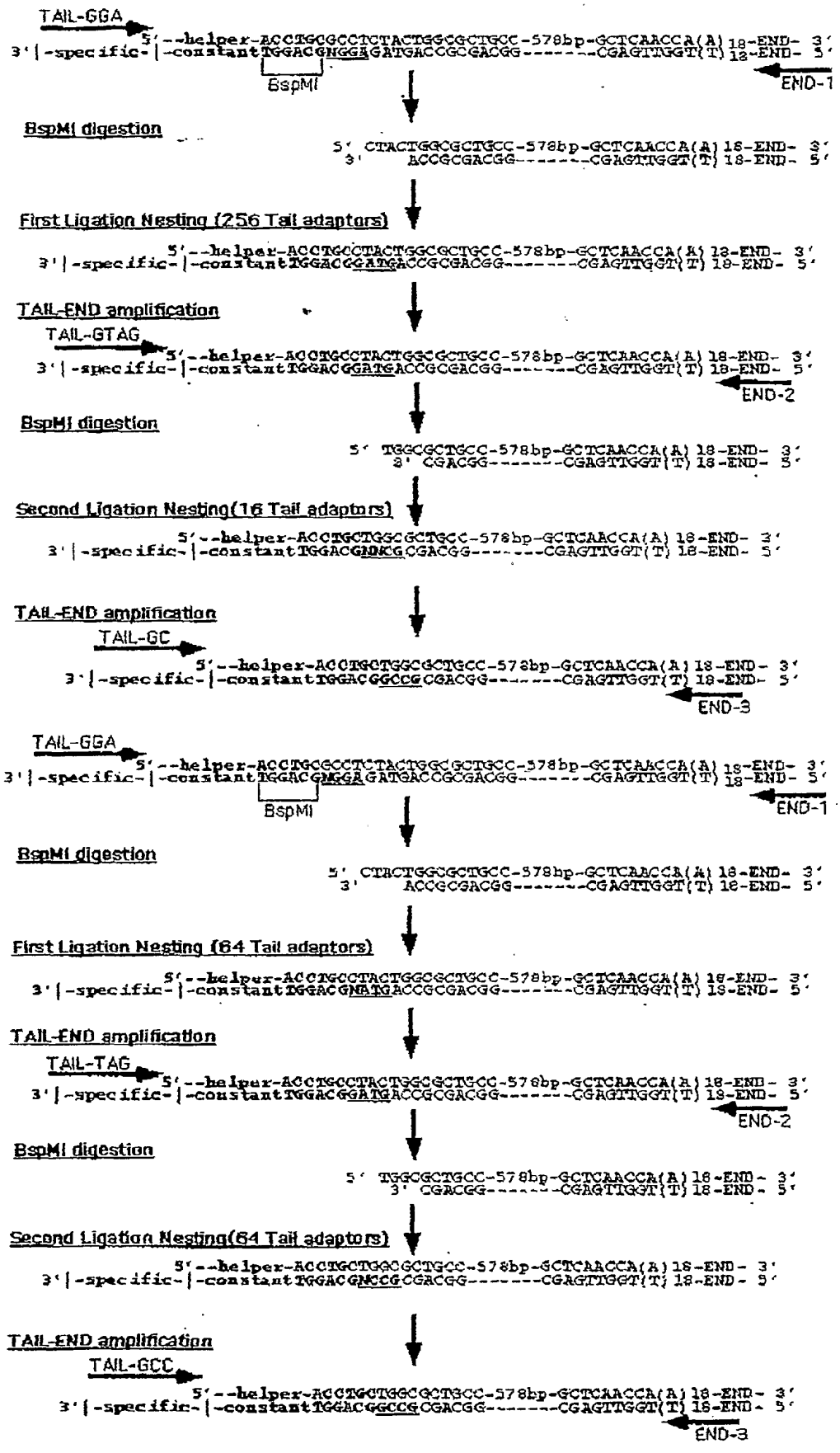

TAIL-GGA
5'--helper-ACCTGCGCCTCTACTGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3'|-specific-|-constantTGGACGNGGAGATGACCGCGACGG-------CGAGTTGGT(T)18-END- 5'
BspMI
END-1
BspMI digestion
5' CTACTGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3' ACCGCGACGG-------CGAGTTGGT(T)18-END- 5'
First Ligation Nesting (256 Tail adaptors)
5'--helper-ACCTGCCTACTGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3'|-specific-|-constantTGGACGGATGACCGCGACGG-------CGAGTTGGT(T)18-END- 5'
TAIL-END amplification
TAIL-GTAG
5'--helper-ACCTGCCTACTGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3'|-specific-|-constantTGGACGGATGACCGCGACGG-------CGAGTTGGT(T)18-END- 5'
END-2
BspMI digestion
5' TGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3' CGACGG-------CGAGTTGGT(T)18-END- 5'
Second Ligation Nesting(16 Tail adaptors)
5'--helper-ACCTGCTGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3'|-specific-|-constantTGGACGNNCGCGACGG-------CGAGTTGGT(T)18-END- 5'
TAIL-END amplification
TAIL-GC
5'--helper-ACCTGCTGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3'|-specific-|-constantTGGACGGCCGCGACGG-------CGAGTTGGT(T)18-END- 5'
END-3

FIG. 2

TAIL-GGA
5'--helper-ACCTGCGCCTCTACTGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3'|-specific-|-constantTGGACGNGGAGATGACCGCGACGG-------CGAGTTGGT(T)18-END- 5'
BspMI
END-1
BspMI digestion
5' CTACTGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3' ACCGCGACGG-------CGAGTTGGT(T)18-END- 5'
First Ligation Nesting (64 Tail adaptors)
5'--helper-ACCTGCCTACTGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3'|-specific-|-constantTGGACGNATGACCGCGACGG-------CGAGTTGGT(T)18-END- 5'
TAIL-END amplification
TAIL-TAG
5'--helper-ACCTGCCTACTGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3'|-specific-|-constantTGGACGGATGACCGCGACGG-------CGAGTTGGT(T)18-END- 5'
END-2
BspMI digestion
5' TGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3' CGACGG-------CGAGTTGGT(T)18-END- 5'
Second Ligation Nesting(64 Tail adaptors)
5'--helper-ACCTGCTGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3'|-specific-|-constantTGGACGNCCGCGACGG-------CGAGTTGGT(T)18-END- 5'
TAIL-END amplification
TAIL-GCC
5'--helper-ACCTGCTGGCGCTGCC-578bp-GCTCAACCA(A)18-END- 3'
3'|-specific-|-constantTGGACGGCCGCGACGG-------CGAGTTGGT(T)18-END- 5'
END-3

FIG. 3

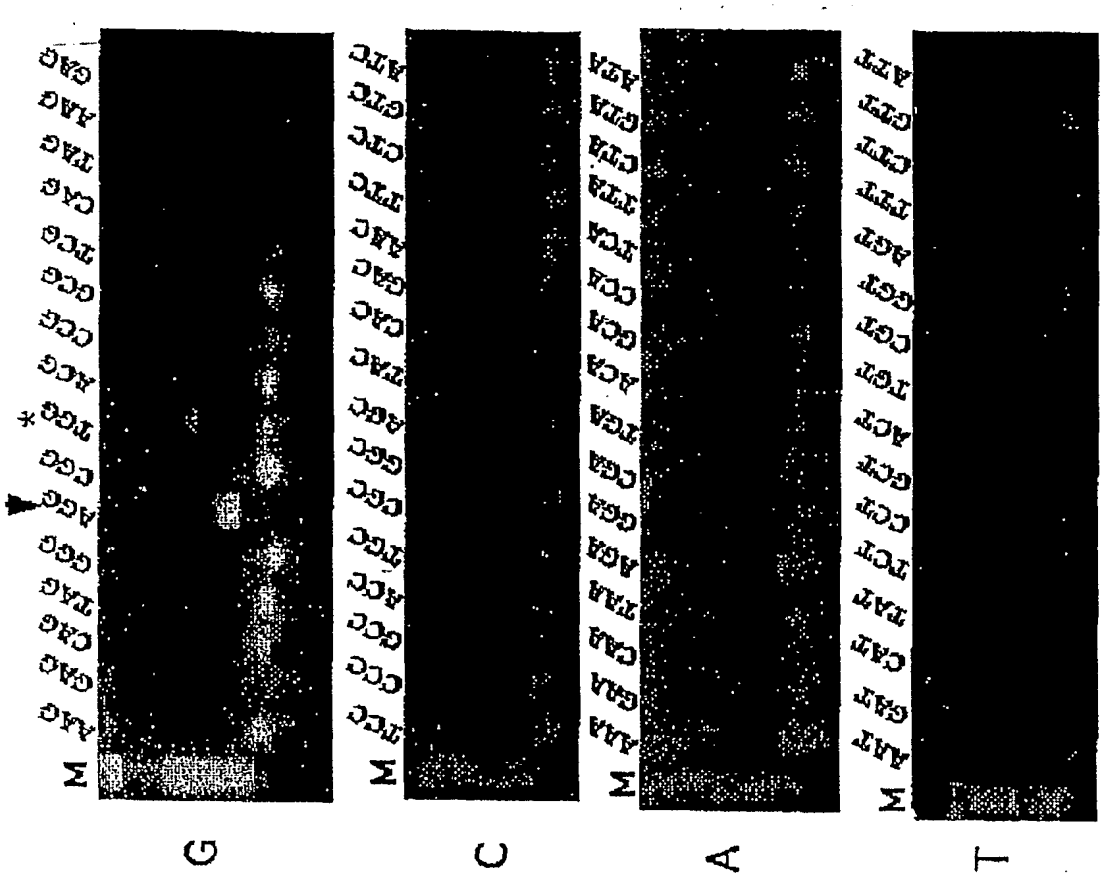

G
C
A
T
M
*

FIG. 5

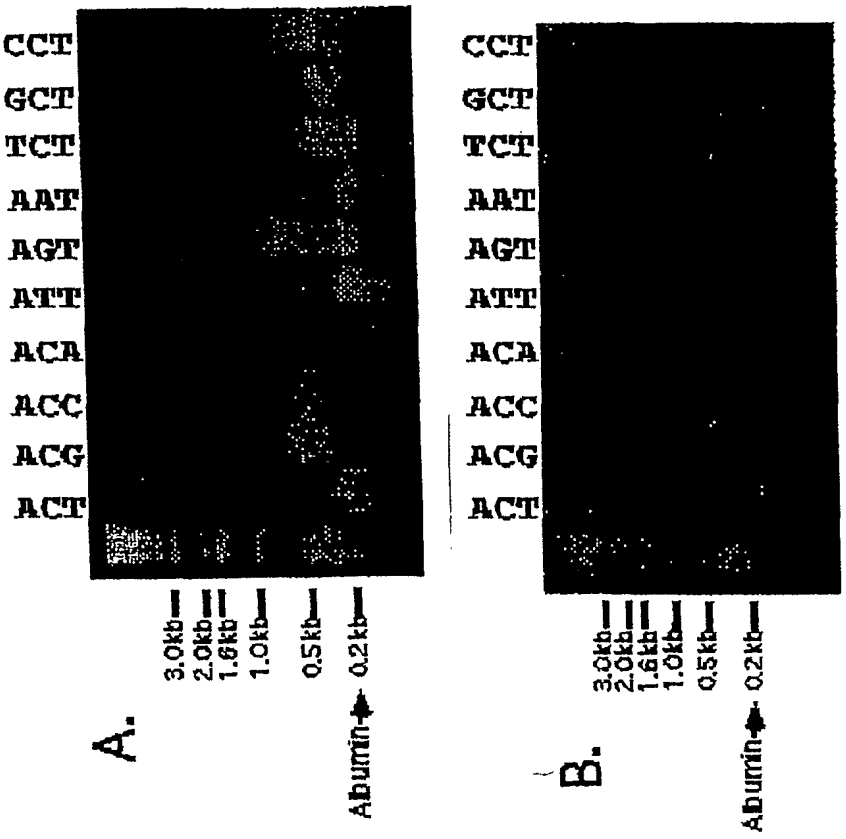

CCT
GCT
TCT
AAT
AGT
ATT
ACA
ACC
ACG
ACT
3.0kb
2.0kb
1.6kb
1.0kb
0.5kb
0.2kb
Albumin
A.
B.

FIG. 4

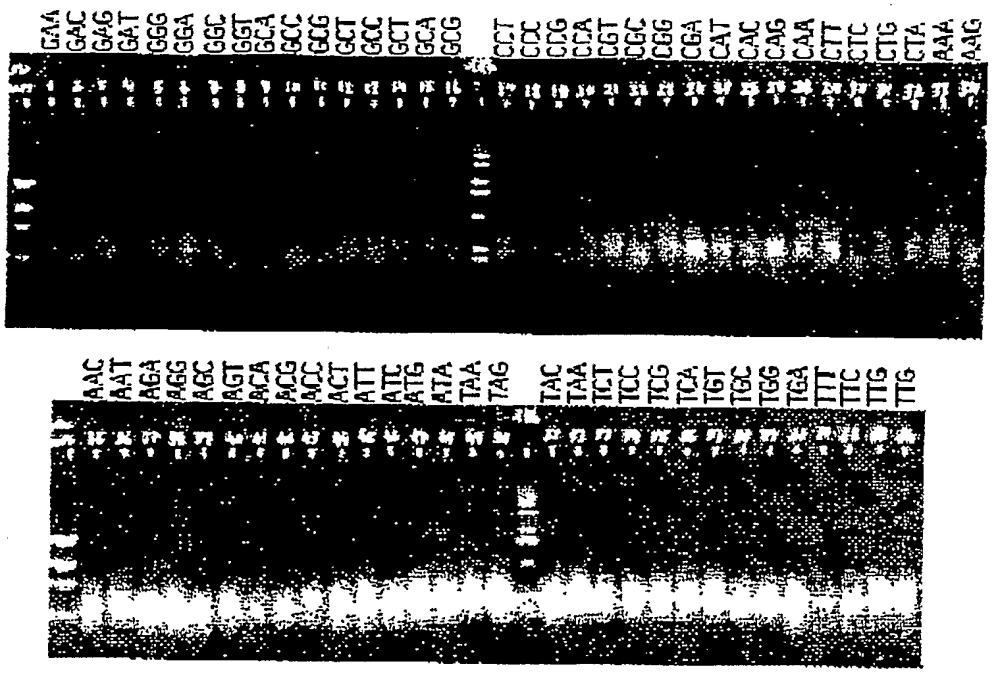

GAA GAC GAG GAT GGG GGA GGC GGT GCA GCC GCG GCT GCC GCT GCA GCG CCT CCC CCG CCA CGT CGC CGG CGA CAT CAC CAG CAA CTT CTC CTG CTA AAA AAG
AAC AAT AGA AGG AGC AGT ACA ACG ACC ACT ATT ATC ATG ATA TAA TAG TAC TAA TCT TCC TCG TCA TGT TGC TGG TGA TTT TTC TTG TTG

FIG. 6

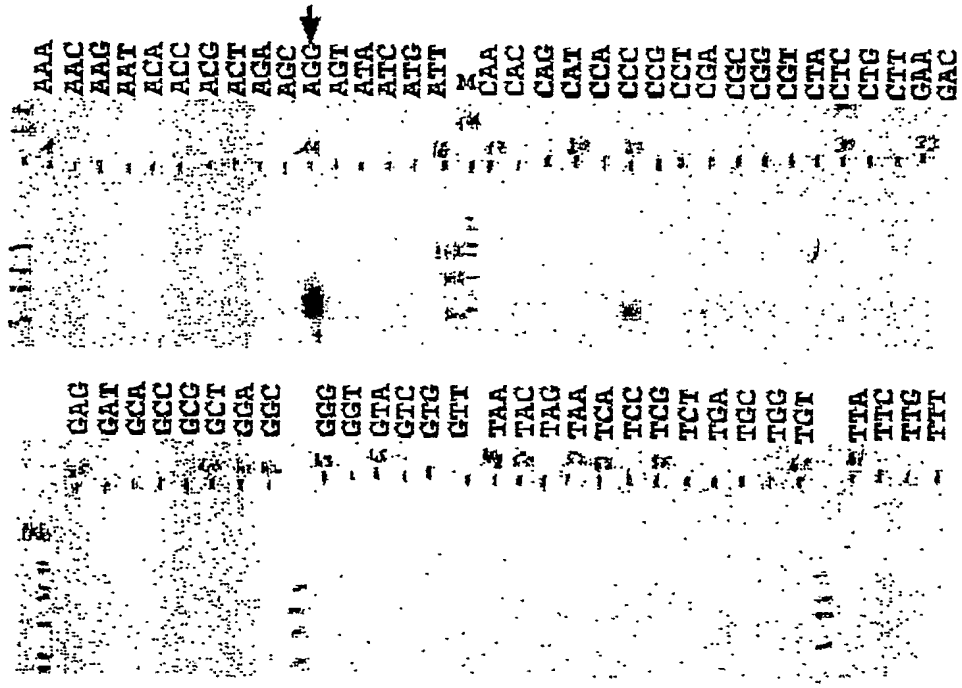

AAA AAC AAG AAT ACA ACC ACG ACT AGA AGC AGG AGT ATA ATC ATG ATT CAA CAC CAG CAT CCA CCC CCG CCT CGA CGC CGG CGT CTA CTC CTG CTT GAA GAC
GAG GAT GCA GCC GCG GCT GGA GGC GGG GGT GTA GTC GTG GTT TAA TAC TAG TAA TCA TCC TCG TCT TGA TGC TGG TGT TTA TTC TTG TTT

FIG. 7

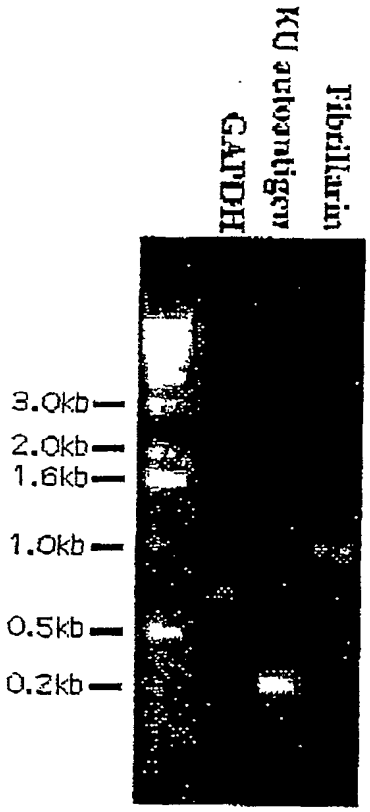

GAPDH
KU autoantigen
Fibrillarin
3.0kb
2.0kb
1.6kb
1.0kb
0.5kb
0.2kb

FIG. 8

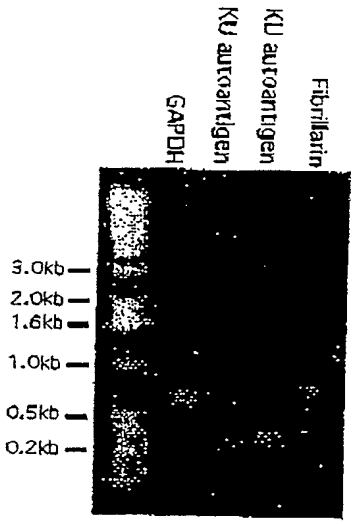

GAPDH
KU autoantigen
KU autoantigen
Fibrillarin
3.0kb
2.0kb
1.6kb
1.0kb
0.5kb
0.2kb

FIG. 9

5′ XXXN GCAGGT ACGTCGTACC GCGGCCGC-x-x-x-x-x-x-x-x-x-x-x-x-3′

Bases 4 BspMI(6)constant(10) NOTI(8) Tail (20)

Primer set J2

5′ TTTNGCAGGTACGTCGTACC GCGGCCGC GTGAGCTTGAGTCGCGTGGA
5′ TTGNGCAGGTACGTCGTACC GCGGCCGC CCAACGTCGCGAGTTAGTCAG
5′ TTCNGCAGGTACGTCGTACC GCGGCCGC AGGTAGACGCGGTATGTTCGTA
5′ TTANGCAGGTACGTCGTACC GCGGCCGC CGGTGCTAGAGTCGCGTGTT
5′ TGTNGCAGGTACGTCGTACC GCGGCCGC CGACAGTACCGCGACAGCTA
5′ TGGNGCAGGTACGTCGTACC GCGGCCGC GCACTTAACTACGCCGACGAAG
5′ TGCNGCAGGTACGTCGTACC GCGGCCGC gTACTAGCCTAACCGAGGCGTA
5′ TGANGCAGGTACGTCGTACC GCGGCCGC TCGGATCACGTACACGTGCT

5′ TCTNGCAGGTACGTCGTACC GCGGCCGC GTACGTCGCCTAGTCGACCTG
5′ TCGNGCAGGTACGTCGTACC GCGGCCGC cTCTCCTAACGGACCGACTAAC
5′ TCCNGCAGGTACGTCGTACC GCGGCCGC CGTTCCGATCTAGCGGTATCTT
5′ TCANGCAGGTACGTCGTACC GCGGCCGC gcACCCGTACaGGATGTGAG
5′ TATNGCAGGTACGTCGTACC GCGGCCGC GCAACGCGCTATGCTCGTag
5′ TAGNGCAGGTACGTCGTACC GCGGCCGC GACTgTGGAACTACGACGATCg
5′ TACNGCAGGTACGTCGTACC GCGGCCGC aGCaGACCGAACCCTAGTCGC
5′ TAANGCAGGTACGTCGTACC GCGGCCGC cATACGTCGTAgggTTCGCGA

5′ GTTNGCAGGTACGTCGTACC GCGGCCGC ctCTCATACGCGTCTGCGCGT
5′ GTGNGCAGGTACGTCGTACC GCGGCCGC gAGTgTGCCTTACGTCGAGttc
5′ GTCNGCAGGTACGTCGTACC GCGGCCGC GTcACGTtGCGGCCTTAGTC
5′ GTANGCAGGTACGTCGTACC GCGGCCGC GagGTACGAgACTTGACACACG
5′ GGTNGCAGGTACGTCGTACC GCGGCCGC GACcAGttGCCTAACGGAcACT
5′ GGGNGCAGGTACGTCGTACC GCGGCCGC GCAACTAGTCTCGACCTGCGA
5′ GGCNGCAGGTACGTCGTACC GCGGCCGC GTACCTCGACGACCGTACTGTg
5′ GGANGCAGGTACGTCGTACC GCGGCCGC ACGCGTGATAGTACGGAGTCG

5′ GCTNGCAGGTACGTCGTACC GCGGCCGC CACTAGAGCGGCGTCAGTCTA
5′ GCGNGCAGGTACGTCGTACC GCGGCCGC GCACAGCGCTAGCACAGGA
5′ GCCNGCAGGTACGTCGTACC GCGGCCGC TACCGACAGTCCTCTGCGTGC
5′ GCANGCAGGTACGTCGTACC GCGGCCGC CTACGCTACGTTGCGAAGAAGGTA
5′ GATNGCAGGTACGTCGTACC GCGGCCGC GTCTGTCGTACCTGTCAGTGACTg
5′ GAGNGCAGGTACGTCGTACC GCGGCCGC ATCGAACCGTGCTCCTTGG
5′ GACNGCAGGTACGTCGTACC GCGGCCGC AGGTTGAGGTGTACGCGATAGC
5′ GAANGCAGGTACGTCGTACC GCGGCCGC GACTTcAACCCCTGACGTACACa

5′ CTTNGCAGGTACGTCGTACC GCGGCCGC CTACTCGCGAGAGAGGGCTATG
5′ CTGNGCAGGTACGTCGTACC GCGGCCGC CTTGATCCGTAGTCGAGACGG
5′ CTCNGCAGGTACGTCGTACC GCGGCCGC GTACAGACGTAGCGATCGCaG
5′ CTANGCAGGTACGTCGTACC GCGGCCGC gTGACTAACGAGGTCTGTAAGCTa
5′ CGTNGCAGGTACGTCGTACC GCGGCCGC GTCTgAGAGTCGACTgCGCTAAG
5′ CGGNGCAGGTACGTCGTACC GCGGCCGC CTcAGTAAGCCGGAGTCTAGCTAg
5′ CGCNGCAGGTACGTCGTACC GCGGCCGC CGCCCTAAACGGGATCGAGCGA
5′ CGANGCAGGTACGTCGTACC GCGGCCGC CGTACAGGCTAGGGGTTAGTCG

5′ CCTNGCAGGTACGTCGTACC GCGGCCGC CGATCGCTCTAGTGCCTACG
5′ CCGNGCAGGTACGTCGTACC GCGGCCGC gAcTGCGATTCGTGACACTAGT
5′ CCCNGCAGGTACGTCGTACC GCGGCCGC TGCGTAATAGCGACTGTACCCt

5′ CCANGCAGGTACGTCGTACC GCGGCCGC cTAGGTCATCCCTCCGGTAC
5′ CATNGCAGGTACGTCGTACC GCGGCCGC gATCGGACTAATCCGCTACGT
5′ CAGNGCAGGTACGTCGTACC GCGGCCGC GACTACCGACTAGTCGTGCGAC
5′ CACNGCAGGTACGTCGTACC GCGGCCGC TAGGGCCCTAACGTAGCTCG
5′ CAANGCAGGTACGTCGTACC GCGGCCGC TACCTAGCCCTAACGGGTCG

5′ ATTNGCAGGTACGTCGTACC GCGGCCGC TAGTGCGCGGTACTACCGACT
5′ ATGNGCAGGTACGTCGTACC GCGGCCGC AGaCGGCTATGCGTCGGGa
5′ ATCNGCAGGTACGTCGTACC GCGGCCGC ACCTACGAACACGCGTAACTCg
5′ ATANGCAGGTACGTCGTACC GCGGCCGC AgcTACGTGGGTGGCAGAC
5′ AGTNGCAGGTACGTCGTACC GCGGCCGC TACCGATACGGTCGACCATC
5′ AGGNGCAGGTACGTCGTACC GCGGCCGC GTACGCTAGGTaggAACTAAGCG
5′ AGCNGCAGGTACGTCGTACC GCGGCCGC CGGACGACTAGTTGCTAGCGTC
5′ AGANGCAGGTACGTCGTACC GCGGCCGC GTGAACCTACGCGTTGACGC

5′ ACTNGCAGGTACGTCGTACC GCGGCCGC GTGTCTCGGGCTAGGCGTAGA
5′ ACGNGCAGGTACGTCGTACC GCGGCCGC TCCGTGGTGTCCATGGGAG
5′ ACCNGCAGGTACGTCGTACC GCGGCCGC CTACGCGTAACGCTAGCAGGT
5′ ACANGCAGGTACGTCGTACC GCGGCCGC gAAGAGCCGTAAGGTACGGCT
5′ AATNGCAGGTACGTCGTACC GCGGCCGC gTACGTCAGCGTACGCTAAGTC
5′ AAGNGCAGGTACGTCGTACC GCGGCCGC TCTAGGTTCCGTTGTAGCGCT
5′ AACNGCAGGTACGTCGTACC GCGGCCGC AGCAACGAGACGACACGAC
5′ AAANGCAGGTACGTCGTACC GCGGCCGC GTCTAGAACCCACGCACGGTA

FIG. 10

Name	Sequence
Tail-TTT 5′	TCCACGCGACTCAAGCTCAC
Tail-TTG 5′	CTGACTAACTCGCGACGTTGG
Tail-TTC 5′	TACGAACATACCGCGTCTACCT
Tail-TTA 5′	AACACGCGACTCTAGCACCG
Tail-TGT 5′	TAGCTGTCGCGGTACTGTCG
Tail-TGG 5′	CTTCGTCGGCGTAGTTAAGTGC
Tail-TGC 5′	TACGCCTCGGTTAGGCTAGTAC
Tail-TGA 5′	AGCACGTGTACGTGATCCGA
Tail-TCT 5′	CAGGTCGACTAGGCGACGTAC
Tail-TCG 5′	GTTAGTCGGTCCGTTAGGAGAG
Tail-TCC 5′	AAGATACCGCTAGATCGGAACG
Tail-TCA 5′	CTCACATCCTGTACGGGTGC
Tail-TAT 5′	CTACGAGCATAGCGCGTTGC
Tail-TAG 5′	CGATCGTCGTAGTTCCACAGTC
Tail-TAC 5′	GCGACTAGGGTTCGGTCTGCT
Tail-TAA 5′	TCGCGAACCCTACGACGTATG
Tail-GTT 5′	ACGCGCAGACGCGTATGAGAG
Tail-GTG 5′	GAACTCGACGTAAGGCACACTC
Tail-GTC 5′	GACTAAGGCCGCAACGTGAC
Tail-GTA 5′	CGTGTGTCAAGTCTCGTACCTC
Tail-GGT 5′	AGTGTCCGTTAGGCAACTGGTC
Tail-GGG 5′	TCGCAGGTCGAGACTAGTTGC
Tail-GGC 5′	CACAGTACGGTCGTCGAGGTAC
Tail-GGA 5′	CGACTCCGTACTATCACGCGT
Tail-GCT 5′	TAGACTGACGCCGCTCTAGTG
Tail-GCG 5′	TCCTGTGCTAGCGCTGTGC
Tail-GCC 5′	GCACGCAGAGGACTGTCGGTA
Tail-GCA 5′	TACCTTCTTCGCAACGTAGCGTAG
Tail-GAT 5′	CAGTCACTGACAGGTACGACAGAC
Tail-GAG 5′	CCAAGGAGCACGGTTCGAT
Tail-GAC 5′	GCTATCGCGTACACCTCAACCT
Tail-GAA 5′	TGTGTACGTCAGGGGTTGAAGTC
Tail-CTT 5′	CATAGCCCTCTCTCGCGAGTAG
Tail-CTG 5′	CCGTCTCGACTACGGATCAAG
Tail-CTC 5′	CTGCGATCGCTACGTCTGTAC
Tail-CTA 5′	TAGCTTACAGACCTCGTTAGTCAC
Tail-CGT 5′	CTTAGCGCAGTCGACTCTCAGAC
Tail-CGG 5′	CTAGCTAGACTCCGGCTTACTGAG
Tail-CGC 5′	TCGCTCGATCCCGTTTAGGGCG
Tail-CGA 5′	CGACTAACCCCTAGCCTGTACG
Tail-CCT 5′	CGTAGGCACTAGAGCGATCG
Tail-CCG 5′	ACTAGTGTCACGAATCGCAGTC
Tail-CCC 5′	AGGGTACAGTCGCTATTACGCA
Tail-CCA 5′	GTACCGGAGGGATGACCTAG
Tail-CAT 5′	ACGTAGCGGATTAGTCCGATC
Tail-CAG 5′	GTCGCACGACTAGTCGGTAGTC
Tail-CAC 5′	CGAGCTACGTTAGGGCCCTA
Tail-CAA 5′	CGACCCGTTAGGGCTAGGTA
Tail-ATT 5′	AGTCGGTAGTACCGCGCACTA
Tail-ATG 5′	TCCCGACGCATAGCCGTCT
Tail-ATC 5′	CGAGTTACGCGTGTTCGTAGGT
Tail-ATA 5′	GTCTGCCACCCACGTAGCT
Tail-AGT 5′	GATGGTCGACCGTATCGGTA
Tail-AGG 5′	CGCTTAGTTCCTACCTAGCGTAC
Tail-AGC 5′	GACGCTAGCAACTAGTCGTCCG
Tail-AGA 5′	GCGTCAACGCGTAGGTTCAC
Tail-ACT 5′	TCTACGCCTAGCCCGAGACAC
Tail-ACG 5′	CTCCCATGGACACCACGGA
Tail-ACC 5′	ACCTGCTAGCGTTACGCGTAG
Tail-ACA 5′	AGCCGTACCTTACGGCTCTTC
Tail-AAT 5′	GACTTAGCGTACGCTGACGTAC
Tail-AAG 5′	AGCGCTACAACGGAACCTAGA
Tail-AAC 5′	GTCGTGTCGTCTCGTTGCT
Tail-AAA 5′	TACCGTGCGTGGGTTCTAGAC

FIG. 11

TnewTTT	5′ AACGACGCGTCGCGGTACCAG	TnewCTT	5′ ACGTCTGTACCCGTCTCGACTAC
TnewTTG	5′ GGGCTCTAGAACTGACTCCAGAC	TnewCTG	5′ GGATCAAGCATAGCCCGTCTC
TnewTTC	5′ GTAGTAGCGGCGACTAGTACC	TnewCTC	5′ TCGCGAGTAGGTGTACGTCAG
TnewTTA	5′ GGCAGGGATCGACCTAGGGTAC	TnewCTA	5′ GGGTTAAGTCGCTATCGCGTAC
TnewTGT	5′ AGTCGCTATTACGCACCTAGTG	TnewCGT	5′ ACCTCAACCTCCAAGGAGCAC
TnewTGG	5′ TCACGAATCGCAATCCCGTAGG	TnewCGG	5′ GGCGTTCGATAGTCACTGACAG
TnewTGC	5′ CACTAGAGCGACGTACCGTGC	TnewCGC	5′ GTACGACAGACTACCTTCTTCG
TnewTGA	5′ GTGGGTTCTAGACCGCGTCGT	TnewCGA	5′ CAACGTAGCGTAGCGTCACGCAG
TnewTCT	5′ GTCGATCTTTGCTAGCGCTAC	TnewCAT	5′ AGGACTGTCGGTATCCTTGTGC
TnewTCG	5′ AACGGAACCTAGAGACTTAGC	TnewCAG	5′ TAGCGCTCGTGCTAGACTGACG
TnewTCC	5′ .GTACGCTGACGTATAGCCGTAC	TnewCAC	5′ CCGCTCTAGTGCGACTCCGT
TnewTCA	5′ CTTACGGCTCTTACCTGCTAG	TnewCAA	5′ ACTATCACGCGTACAGTACGG
TnewTAT	5′ CGTTACGCGTAGACTCCCATG		
TnewTAG	5′ GACACCACGGGATCTACGCCT	TnewATT	5′ TCGTCGAGGTACTTCGCAGGT
TnewTAC	5′ AGCCCGAGACACAGCGTCAAC	TnewATG	5′ CGAGACTAGTTGCAGTTCCGT
TnewTAA	5′ GCGTAGGTTCACGACGCTAGC	TnewATC	5′ TAGGCaaCTAGTCCGTGTGTC
		TnewATA	5′ AAGTTCGTACCtCGACTAAGC
TnewGTT	5′ AACTAGTCGTCCGCGCTTAGT	TnewAGT	5′ GCCGCGACGTAACGaaCTCGA
TnewGTG	5′ TTACCTAGCGTACGGATGGTC	TnewAGG	5′ CGTAAGGCATACTACGCGCAG
TnewGTC	5′ GACCGTATCGGTAGTCTGCCA	TnewAGC	5′ ACGCGTATGAGaAGTCGCGAAc
TnewGTA	5′ CCGCACGTAACGTGAGTTACG	TnewAGA	5′ ccTACGACGTATCGCGACTAG
TnewGGT	5′ CGTGTTCGTAGGTCCGGCGAC		
TnewGGG	5′ GCATAGGCCGCCTAGTCGGTAG	TnewACT	5′ GGTTCGGTCGCtcGATCGTCG
TnewGGC	5′ TACCGCGCACTAGCGACCCGT	TnewACG	5′ TAGTTCCAAGTCGCGACGAGC
TnewGGA	5′ TAGGGCTAGGTACGAGCTACG	TnewACC	5′ .ATAGCGCGTTGCCCTCACATC
		TnewACA	5′ CtGTACGGGTgcAAGATACCG
TnewGCT	5′ TTAGGGCCCTAGTCGCACGAC	TnewAAT	5′ CTAGATCGGAACGGTTAGTCG
TnewGCG	5′ TAGTCGGTAGTCACCCGTAGC	TnewAAG	5′ GTCCGTTAGGAGACAGGTCGA
TnewGCC	5′ GGATTAGTCCGATGCGACTAACC	TnewAAC	5′ CTAGGCGACGTACCAGCACGT
TnewGCA	5′ CCTAGCCTGTACGTCGCTCGATCC	TnewAAA	5′ GTACGTGATCCGATACGCCTC
TnewGAT	5′ CGTTTAGGGCGTAGCTAGACTCC		
TnewGAG	5′ GGCTTACTAGCTTAGCGAGTC		
TnewGAC	5′ GACTCTAGACAGCTTACAGACC		
TnewGAA	5′ TCGTTAGTCACtGCGATCGCT		

FIG. 12

tailTTTT	5'AACGCAGTGTTCGTTCGACGA
tailTTTG	5'TGCAGAGCGGAACGAGaCGTA
tailTTTC	5'AGCGCACGTCGTCTAGCGAAG
tailTTTA	5'TCTGAGACGGaGTACGAGCGA
tailTTGT	5'ACAGTGACCGTTTTCGCGCAT
tailTTGG	5'AcTCTGGGACGACGAAAAGCG
tailTTGC	5'TGACCGAACCGGGTTTACCAG
tailTTGA	5'ACGAGACGTCTCGGACTATCG
tailTTCT	5'CAACAACGTGCCGTTCGATAG
tailTTCG	5'GTACCACCCGAACGGTCGTAG
tailTTCC	5'GCACAACCGTCGACCGTACGA
tailTTCA	5'AAGCCGAGACGAGGTCTAACG
tailTTAT	5'ATCGCTGCGATCGGACGTTAG
tailTTAG	5'ACCGCAGACGTTCCGATACCG
tailTTAC	5'TCTACGTACGACGGTTCGGTA
tailTTAA	5'TACACCACGTGAATCCGCTAG
tailTGTT	5'ATCCTGGACAGAGTCGTCGAC
tailTGTG	5'TCGTGAGTCAAGAACCGTCGA
tailTGTC	5'ACAGCACACGTGATCcTTACG
tailTGTA	5'TGGTACACGCTCGATCCGTAAG
tailTGGT	5'CTCACTCGGGTCGTTGCGTATG
tailTGGG	5'gGaTTACACACGCAAgGATACG
tailTGGC	5'TGGCATCGTGCTTCTTCCGAT
tailTGGA	5'GACGTCCTCGCGAGAAATCGG
tailTGCT	5'AGTATCCAGCAGTGGGATGCG
tailTGCG	5'ACGAAGAGCGACCGAACCGTA
tailTGCC	5'GCAACTGCGGTTCGACGAATG
tailTGCA	5'ACGTTCGCGAGTCGAAATTCG
tailTGAT	5'AACGTGTCACTGCGTCGCGTT
tailTGAG	5'GTCTAGACGGAGAAGCAAAGC
tailTGAC	5'CGTTAGCGCTCGACGTTACGT
tailTGAA	5'GATCACTCCGCACGTCACGTA
tailTCTT	5'ACTAGTTACCGAGCGTCTACG
tailTCTG	5'CTATGCGAGAGACGCTCGTAG
tailTCTC	5'ACACGAACGGATGCGTTTCGC
tailTCTA	5'TACTAGCAGCAACGAAGCGAA
tailTCGT	5'CTAGACTCCGGTGTCGATCGT
tailTCGG	5'CGACTACGTCCCGACAACGAT
tailTCGC	5'AACTCGGAAGACGATGGTCGT
tailTCGA	5'AAGTATGGACGCATCGACGAC
tailTCCT	5'TGAAGGTCGACACGTTCGGTT
tailTCCG	5'AATACCGCGCAAACGTAACCA
tailTCCC	5'TAGACACAGGACCAGGGTTCG
tailTCCA	5'AGTACTTCGTGACGAGCGAAC
tailTCAT	5'AACTAGAAGCTGCGGTTTGCG
tailTCAG	5'ACTAGCTGCGAACGGTCGCAA
tailTCAC	5'AGCATACGCTTACCTGCGACT
tailTCAA	5'ACGTGGAGCCTACGATAGTCG
tailTATT	5'CCTAACCTCGAATCGCTCGAT
tailTATG	5'ACCACGGCGCTACGGTATCGA
tailTATC	5'ATGCCGTCGAGAGAGTTCGGT
tailTATA	5'TCAACCACGAGTGACGATCGA
tailTAGT	5'ACTATCCTCGTCGTCAGTCGC
tailTAGG	5'AGGTTATCCGTCTGCCACGAC
tailTAGC	5'TCCACGACTGACGAACCGCAT
tailTAGA	5'GAGCTAGACGGAATCGATACG
tailTACT	5'AACGGAGCCGTCGATCTTCGT
tailTACG	5'CAGTACGTGGTCTTCGTTCGA
tailTACC	5'CGATCACCGCCGAAGTCAGCA
tailTACA	5'GTCAGACTCGCGTCTACGAAC
tailTAAT	5'CCATTCGAGTAAACGCGATTG
tailTAAG	5'ATAGTCGCTCGTTCCGAATCG
tailTAAC	5'GCCTTAGAGCCAGGAAGAACG
tailTAAA	5'GGTTCACGCACGTTAGCGTTC
tailGTTT	5'CCAATTCCTTCCCTGGCTCATC
tailGTTG	5'TCTCGGTCGCCTCGTCTAATC
tailGTTC	5'AGACTCCTCAGCTGACCTAGTC
tailGTTA	5'AGTCAGCTCGCCACTCGTAGT
tailGTGT	5'AGAGTACTCGAGTCAGTAGGC
tailGTGG	5'ACAGAGGAGTCGGGAACAACG
tailGTGC	5'CTTGGGTACCTGTGTCCGTTG
tailGTGA	5'ACAGTACGAAGCAATCTGTGA
tailGTCT	5'AGCTCGGAGAGCATAAGGACG
tailGTCG	5'TCTCGGGCATTACTGGATAGG
tailGTCC	5'CCTTAACCtGATCTGTCcCATG
tailGTCA	5'GTGCGAGTCCAGTTTGACTGA
tailGTAT	5'GGTGGCCAACCACAGCCTTC
tailGTAG	5'TGAGATGAGGTGTACGACTGC
tailGTAC	5'TGTCAATGCGCCAGTTGTCTA
tailGTAA	5'GCACCAACACCTAGTGGCATC
tailGGTT	5'GATCTGTAGAGCGGGAGGTCT
tailGGTG	5'TGGCTAAGGGTGCTGCCACGC
tailGGTC	5'ATGAGACTCCAGCCGAAACCT
tailGGTA	5'AGTGTAGGGACGACCTGCAGA
tailGGGT	5'GGCAACGGCATAGCTGATACA
tailGGGG	5'GATGCTGAGGTATGAGGCAACG
tailGGGC	5'ACGTCATTTGGCCTGTCTGCT
tailGGGA	5'GACTCACGTGCTCGAACTGCT
tailGGCT	5'AGTCGGCaTGTgGCAcAtcTc
tailGGCG	5'ACTCGGTAGACAGCCGCTAAC
tailGGCC	5'CTGGGACACGGTCACTATTCAC
tailGGCA	5'ACCCTTGGAACGCTGTACACA
tailGGAT	5'TCCGGACACGTAGTGAGACGT
tailGGAG	5'TGCCTTGCACTCTTACCTAGC
tailGGAC	5'TAGCCAGTATCGTGCACTTGG
tailGGAA	5'AAGCTTACCACCCTACACGAA
tailGCTT	5'AgGATGaTGACaTGGgTCGAa
tailGCTG	5'AACCTCCATGACAAGTCCTCC
tailGCTC	5'AACACCGTGGGACAGACATCT
tailGCTA	5'CCACGGAACATACAGGGCATT
tailGCGT	5'CATGAGCGTGGAGCTAAGCAT
tailGCGG	5'CATCTGTCACAAGGTACGAGG
tailGCGC	5'AgGaGATgGAaCGCTCGcACA
tailGCGA	5'TCTGTGTCCTCGACCAGCATC
tailGCCT	5'AACTCCAGGTGGAAGCTGGTT

FIG. 13

tailGCCG	5'CAGACTCACATCGAACGTCAC
tailGCCC	5'TGTAACTCCGAATGGGACACC
tailGCCA	5'GTTGATGCTCTCCCTCACCTG
tailGCAT	5'GAGTCTGCCAACAAGGTCGAG
tailGCAG	5'GTTGTGAGGAACCGCAATGCA
tailGCAC	5'ACCTCAGTGAACAGCTCTCAG
tailGCAA	5'GATCCAGGTCGCTATCCACTG
tailGATT	5'CcACATGCGaTCTCAAaTCCa
tailGATG	5'TTGTCGTGACGACCTAGACGC
tailGATC	5'TTGAGGCGTCTAATCATCGGG
tailGATA	5'CGCTCAGCAATCGCCACTATC
tailGAGT	5'CATTATCACACATGAGCCGCC
tailGAGG	5'GAGGGGCAAGAGAAAACCACC
tailGAGC	5'AAGTCCAGCGAGCTGTCTTCC
tailGAGA	5'AgGCCgctTCTCAGtAAGGTC
tailGACT	5'GTGTACGCAGAGAACCCCACA
tailGACG	5'GGTCTCCTGGACAACAGTTCC
tailGACC	5'CaGttGCATCACtCtggCATC
tailGACA	5'AAGACCGAATCGCGAAATGAG
tailGAAT	5'GTTCAGACCACCCGGTTCACA
tailGAAG	5'TGCTACAGCAGGATCCTCTGG
tailGAAC	5'GATACCTAGACCGGCAGCAAC
tailGAAA	5'CACTGAGAGCTAGGAAACCCAC
tailCTTT	5'GGGATAAATCCTGATGCCGTC
tailCTTg	5'CAGTCTCAACCCTTGCCTGTC
tailCTTc	5'TCACGGAGCTCACCTAAGCAC
tailCTTa	5'GATTTGGAGCTGACCTGATGC
tailCTGT	5'GATGTATCTATGAAATCGAGT
tailCTGg	5'CAACCCCGTAACTCCGTTCAG
tailCTGc	5'CGTCGACTTGTGCGACCTTCG
tailCTGa	5'AACACGCACAACCAGGTCATG
tailCTCT	5'TCGTCTCCAGCTACTGGACTC
tailCTCg	5'TACGCTCAACACTTACAGACG
tailCTCc	5'GGGCAACAGCACCTACTATAC
tailCTCa	5'CGTCTGACCAGTCTTCCACTC
tailCTAT	5'GGGAGAGGTGTTTTCCAGTCG
tailCTAg	5'GACCCAAGTAGTCGTCGCGAA
tailCTAc	5'CACCATGGTGAATCAGGCTCC
tailCTAa	5'ACCTGAGTGTGGGAAGGTCGA
tailCGTT	5'TGCGAAACTGTCTGTCGGAAG
tailCGTg	5'GCTTTGGCAATCCTCAAGCAG
tailCGTc	5'TCGCTCCTGACTCATCGAACA
tailCGTa	5'CAGAGTCGGTACCATCTCGAC
tailCGGT	5'GCGGACAAAGGATATGTTGATC
tailCGGg	5'CACTAGGACCTTTTGTCGGAAG
tailCGGc	5'TAAGAGCGGTGCTAGCGTGAG
tailCGGa	5'GGAGCCTCGAGATTCGTTGGT
tailCGCT	5'GCCTGGTCTTTCAGCATGGAC
tailCGCg	5'CTTGTCAGCCGAACGTCTGTC
tailCGCc	5'ACGCTGCAAGGCGGATAACAG
tailCGCa	5'CAGCACATAGACAGGTGCCTCA
tailCGAT	5'ATCATCACGTTGCACCAAGGG
tailCGAg	5'TCCAGAGGAACGTACGACCCT
tailCGAc	5'GAACAGGAGACAGAGCGAGCA
tailCGAa	5'CTACGGTCAGTACGACGTGGA
tailCCTT	5'AAATTATTCGCTGGAGCGCTG
tailCCTg	5'CAGCTGCGGTGTAGCATACAG
tailCCTc	5'ACTCGTAATCGTTCCAGACGC
tailCCTa	5'ATACGTGTTATGGCCGGAAAG
tailCcgT	5'GCTCCGAAGTTAGGTTGGGAA
tailCcgg	5'GTTCACCCTTGCAACGATAGC
tailCcgc	5'AGGGAGACTCCCTACTCGGAT
tailCcga	5'GAGTTGCCAGACATGTACCAG
tailCcCt	5'GCCAGTTTCTTCCCACAAGCA
tailCcCg	5'GTGAACGAGTATGCGACCCAG
tailCcCc	5'TTGCCTGTATTGCAACGCCTA
tailCcCa	5'TGAGCTGCTGGAAGATCAGGA
tailCcAt	5'AGTAGGGGAATACGCAACATGA
tailCcAg	5'GATCCACTTCGAGGAGTGACC
tailCcAc	5'GTACCACATTCGCTCGACACG
tailCcAa	5'CATTTCCCTCTCGAATTGGCA
tailCaTT	5'TCCGATGTATCGCCGAGATGT
tailCaTg	5'ACCAACTGAGAAGGAAGGTCA
tailCaTc	5'CGAATCCTAGTCACCAGTACTC
tailCaTa	5'GGAAGGATGCACTCCTACCGA
tailCagT	5'AATAGCTCCCTCCCTCACCAC
tailCagg	5'GAGGACCATCTGCTACATCTC
tailCagc	5'ATTACTTCGCGGGTCCTAATC
tailCaga	5'CAGCGACAACAAAAGGCTATG
tailCaCT	5'GCGTTGACACCTCATCACTAG
tailCaCg	5'TCTACCACTCACCGTCCGAAC
tailCaCc	5'AGCATGCTTCTGAGGAAGTGC
tailCaCa	5'AGTCATCGTGGCTTGTGTTACA
tailCaAT	5'GACACTTGGCTATGGGTCCCA
tailCaAg	5'CACAGTACGTGAGAGCTCCAA
tailCaAc	5'GAAGCAACCCAACAGGACCAG
tailCaAa	5'AGAGACTCACCAGGAAGCAGCA
tailATTT	5'TGTGGTACAGCAGAAGGCTGA
tailATTG	5'TCCAAGTTCGCCAAAGCAGGA
tailATTC	5'CGTGCGATTCTGGAATGCTTC
tailATTA	5'ACTCGGAATGGTGGGAGAGGA
tailATGT	5'AGCAGATTCTCGAGGAAACCA
tailATGG	5'ACCTCTCTGGTCTGGTCAGCA
tailATGC	5'TGACAAGTGGATGAGTGAGCAG
tailATGA	5'GGATTTTTCGACCGTGGTACA
tailATCT	5'GCCTGAGAGCTTTACTCACCA
tailATCG	5'GCTTAGCTTCTGCGATGGCAC
tailATCC	5'CAGCAGTGTCAGGTAGCCTCA
tailATCA	5'AGACAAGAGGTTCTGGCACCA
tailATAT	5'TGGTGGGTCTATCAAGTCGCA
tailATAG	5'TGTCGTAGCCACTGATGCTAC
tailATAC	5'TCATCCCTGGCATCGATGCTC
tailATAA	5'GAGGTGCCTTCCCAGACAGAG
tailAGTT	5'CGTCTCTGGAGTCGTCCTCTC
tailAGTG	5'TGGAGTCACGGTCTATGGATG
tailAGTC	5'AGTCTCCTGGAATGACGTGGAC
tailAGTA	5'CCAGTGTCCTCACCTAGATCG
tailAGGT	5'AGCCTACGCCAGTTGTCCTTC

tailAGGG	5'CCTTGTAGAGGATACGAACGAC
tailAGGC	5'AGGTAGCACAGCCAGGAACTC
tailAGGA	5'TCGTACACGATCCATCAGCAG
tailAGCT	5'GAACCCTCTGCCTTCGAACAC
tailAGCG	5'CTCAACCTAGACCCCTTAAACC
tailAGCC	5'CTTAGCAACGTCCCAGAGGAG
tailAGCA	5'AGGAGATCACTGCGTCTGCTG
tailAGAT	5'CCAGCTGCTCACTTCATGCTC
tailAGAG	5'ACCAGTCTCTACTGAGGCCAG
tailAGAC	5'CTATTGCACTAGTGCCTGCCA
tailAGAA	5'TGCGGACACGACAGGATGTAG
tailACTT	5'CCAGTGCTACCTCAGATCCGT
tailACTG	5'GAATCGAGCTGAGGCTTCTCA
tailACTC	5'CAGGCGAATTAACCTCAAACG
tailACTA	5'GCTCGGGTATTTGCAGTAGCA
tailACGT	5'TGAGGAGTTACGTGCAGACGA
tailACGG	5'TGACAGTCGCTTGAACCATCC
tailACGC	5'ACAGACCACCAGCTGAGAGTG
tailACGA	5'GTCCATTCCCATCAACCAAGC
tailACCT	5'GTACGTCTAGTCTTGCTTGCAG
tailACCG	5'GACACTTGGGAGCTTCATGGA
tailACCC	5'CCTGCGTTTAACCAATGTGCA
tailACCA	5'ATCTACCTGCAATGATCTGCA
tailACAT	5'AGACCGTCTTCCAGTCGTGCT
tailACAG	5'ACCACCGATGATGTTCATGCT
tailACAC	5'TCCACCACAGTCCAGACTCCA
tailACAA	5'GACGAGTCGACGAGGTGTAAG
tailAATT	5'GACCTACGGAAGCTTAGCCCT
tailAATG	5'ACACCACCGCAACTAGCCAAC
tailAATC	5'CGTTGTGCCTAAGACCTGCGA
tailAATA	5'GGAACCAGAATCGGACCTGAC
tailAAGT	5'TGGAGTTGATGGGTCGAGCTG
tailAAGG	5'GACAGCTATGTTGCCGGTAGC
tailAAGC	5'TCAGAGTGGCACATACTGAGGA
tailAAGA	5'GATGGCACGTAGGCAAGCAAC
tailAACT	5'CTCTGTGCTTCGGGCCTAGTC
tailAACG	5'CGTATCACCTGTGTCCAGCAA
tailAACC	5'CTAACAACGGTGGCGTTCCA
tailAACA	5'TGCAACCTCGATCCCATACG
tailAAAT	5'GTGAGGAGCTGATGAGACTGA
tailAAAG	5'CGAACGGTTACGTCACCAAGG
tailAAAC	5'ACTTCAGTTCCTAGGCTCGTC
tailAAAA	5'AGGTCTCCATCACGACTCCAC

5' GGTACGACGTTCAGCTNGGGIIIAA
5' GGTACGACGTTCAGCTNGGGIIIAC
5' GGTACGACGTTCAGCTNGGGIIIAG
5' GGTACGACGTTCAGCTNGGGIIIAT
5' GGTACGACGTTCAGCTNGGGIIICA
5' GGTACGACGTTCAGCTNGGGIIICC
5' GGTACGACGTTCAGCTNGGGIIICG
5' GGTACGACGTTCAGCTNGGGIIICT

5' GGTACGACGTTCAGCTNGGGIIIGA
5' GGTACGACGTTCAGCTNGGGIIIGC
5' GGTACGACGTTCAGCTNGGGIIIGG
5' GGTACGACGTTCAGCTNGGGIIIGT
5' GGTACGACGTTCAGCTNGGGIIITA
5' GGTACGACGTTCAGCTNGGGIIITC
5' GGTACGACGTTCAGCTNGGGIIITG
5' GGTACGACGTTCAGCTNGGGIIITT

FIG. 17

5′ GGTACGACGTTCAGCTNIIIAAAA
5′ GGTACGACGTTCAGCTNIIIAAAC
5′ GGTACGACGTTCAGCTNIIIAAAG
5′ GGTACGACGTTCAGCTNIIIAAAT
5′ GGTACGACGTTCAGCTNIIIAACA
5′ GGTACGACGTTCAGCTNIIIAACC
5′ GGTACGACGTTCAGCTNIIIAACG
5′ GGTACGACGTTCAGCTNIIIAACT

5′ GGTACGACGTTCAGCTNIIIAAGA
5′ GGTACGACGTTCAGCTNIIIAAGC
5′ GGTACGACGTTCAGCTNIIIAAGG
5′ GGTACGACGTTCAGCTNIIIAAGT
5′ GGTACGACGTTCAGCTNIIIAATA
5′ GGTACGACGTTCAGCTNIIIAATC
5′ GGTACGACGTTCAGCTNIIIAATG
5′ GGTACGACGTTCAGCTNIIIAATT

5′ GGTACGACGTTCAGCTNIIIACAA
5′ GGTACGACGTTCAGCTNIIIACAC
5′ GGTACGACGTTCAGCTNIIIACAG
5′ GGTACGACGTTCAGCTNIIIACAT
5′ GGTACGACGTTCAGCTNIIIACCA
5′ GGTACGACGTTCAGCTNIIIACCC
5′ GGTACGACGTTCAGCTNIIIACCG
5′ GGTACGACGTTCAGCTNIIIACCT

5′ GGTACGACGTTCAGCTNIIIACGA
5′ GGTACGACGTTCAGCTNIIIACGC
5′ GGTACGACGTTCAGCTNIIIACGG
5′ GGTACGACGTTCAGCTNIIIACGT
5′ GGTACGACGTTCAGCTNIIIACTA
5′ GGTACGACGTTCAGCTNIIIACTC
5′ GGTACGACGTTCAGCTNIIIACTG
5′ GGTACGACGTTCAGCTNIIIACTT

5′ GGTACGACGTTCAGCTNIIIAGAA
5′ GGTACGACGTTCAGCTNIIIAGAC
5′ GGTACGACGTTCAGCTNIIIAGAG
5′ GGTACGACGTTCAGCTNIIIAGAT
5′ GGTACGACGTTCAGCTNIIIAGCA
5′ GGTACGACGTTCAGCTNIIIAGCC
5′ GGTACGACGTTCAGCTNIIIAGCG
5′ GGTACGACGTTCAGCTNIIIAGCT

5′ GGTACGACGTTCAGCTNIIIAGGA
5′ GGTACGACGTTCAGCTNIIIAGGC
5′ GGTACGACGTTCAGCTNIIIAGGG
5′ GGTACGACGTTCAGCTNIIIAGGT

5′ GGTACGACGTTCAGCTNIIIAGTA
5′ GGTACGACGTTCAGCTNIIIACTC
5′ GGTACGACGTTCAGCTNIIIAGTG
5′ GGTACGACGTTCAGCTNIIIAGTT

5′ GGTACGACGTTCAGCTNIIIATAA
5′ GGTACGACGTTCAGCTNIIIATAC
5′ GGTACGACGTTCAGCTNIIIATAG
5′ GGTACGACGTTCAGCTNIIIATAT
5′ GGTACGACGTTCAGCTNIIIATCA
5′ GGTACGACGTTCAGCTNIIIATCC
5′ GGTACGACGTTCAGCTNIIIATCG
5′ GGTACGACGTTCAGCTNIIIATCT

5′ GGTACGACGTTCAGCTNIIIATGA
5′ GGTACGACGTTCAGCTNIIIATGC
5′ GGTACGACGTTCAGCTNIIIATGG
5′ GGTACGACGTTCAGCTNIIIATGT
5′ GGTACGACGTTCAGCTNIIIATTA
5′ GGTACGACGTTCAGCTNIIIATTC
5′ GGTACGACGTTCAGCTNIIIATTG
5′ GGTACGACGTTCAGCTNIIIATTT

5′ GGTACGACGTTCAGCTNIIICAAA
5′ GGTACGACGTTCAGCTNIIICAAC
5′ GGTACGACGTTCAGCTNIIICAAG
5′ GGTACGACGTTCAGCTNIIICAAT
5′ GGTACGACGTTCAGCTNIIICACA
5′ GGTACGACGTTCAGCTNIIICACC
5′ GGTACGACGTTCAGCTNIIICACG
5′ GGTACGACGTTCAGCTNIIICACT

5′ GGTACGACGTTCAGCTNIIICAGA
5′ GGTACGACGTTCAGCTNIIICAGC
5′ GGTACGACGTTCAGCTNIIICAGG
5′ GGTACGACGTTCAGCTNIIICAGT
5′ GGTACGACGTTCAGCTNIIICATA
5′ GGTACGACGTTCAGCTNIIICATC
5′ GGTACGACGTTCAGCTNIIICATG
5′ GGTACGACGTTCAGCTNIIICATT

FIG. 14

5′ GGTACGACGTTCAGCTNIIICCAA
5′ GGTACGACGTTCAGCTNIIICCAC
5′ GGTACGACGTTCAGCTNIIICCAG
5′ GGTACGACGTTCAGCTNIIICCAT
5′ GGTACGACGTTCAGCTNIIICCCA
5′ GGTACGACGTTCAGCTNIIICCCC
5′ GGTACGACGTTCAGCTNIIICCCG
5′ GGTACGACGTTCAGCTNIIICCCT

5′ GGTACGACGTTCAGCTNIIICCGA
5′ GGTACGACGTTCAGCTNIIICCGC
5′ GGTACGACGTTCAGCTNIIICCGG
5′ GGTACGACGTTCAGCTNIIICCGT
5′ GGTACGACGTTCAGCTNIIICCTA
5′ GGTACGACGTTCAGCTNIIICCTC
5′ GGTACGACGTTCAGCTNIIICCTG
5′ GGTACGACGTTCAGCTNIIICCTT

5′ GGTACGACGTTCAGCTNIIICGAA
5′ GGTACGACGTTCAGCTNIIICGAC
5′ GGTACGACGTTCAGCTNIIICGAG
5′ GGTACGACGTTCAGCTNIIICGAT
5′ GGTACGACGTTCAGCTNIIICGCA
5′ GGTACGACGTTCAGCTNIIICGCC
5′ GGTACGACGTTCAGCTNIIICGCG
5′ GGTACGACGTTCAGCTNIIICGCT

5′ GGTACGACGTTCAGCTNIIICGGA
5′ GGTACGACGTTCAGCTNIIICGGC
5′ GGTACGACGTTCAGCTNIIICGGG
5′ GGTACGACGTTCAGCTNIIICGGT
5′ GGTACGACGTTCAGCTNIIICGTA
5′ GGTACGACGTTCAGCTNIIICCTC
5′ GGTACGACGTTCAGCTNIIICGTG
5′ GGTACGACGTTCAGCTNIIICGTT

5′ GGTACGACGTTCAGCTNIIICTAA
5′ GGTACGACGTTCAGCTNIIICTAC
5′ GGTACGACGTTCAGCTNIIICTAG
5′ GGTACGACGTTCAGCTNIIICTAT
5′ GGTACGACGTTCAGCTNIIICTCA
5′ GGTACGACGTTCAGCTNIIICTCC
5′ GGTACGACGTTCAGCTNIIICTCG
5′ GGTACGACGTTCAGCTNIIICTCT

5′ GGTACGACGTTCAGCTNIIICTGA
5′ GGTACGACGTTCAGCTNIIICTGC
5′ GGTACGACGTTCAGCTNIIICTGG
5′ GGTACGACGTTCAGCTNIIICTGT
5′ GGTACGACGTTCAGCTNIIICTTA
5′ GGTACGACGTTCAGCTNIIICTTC
5′ GGTACGACGTTCAGCTNIIICTTG
5′ GGTACGACGTTCAGCTNIIICTTT

5′ GGTACGACGTTCAGCTNIIIGAAA
5′ GGTACGACGTTCAGCTNIIIGAAC

5′ GGTACGACGTTCAGCTNIIIGAAG
5′ GGTACGACGTTCAGCTNIIIGAAT
5′ GGTACGACGTTCAGCTNIIIGACA
5′ GGTACGACGTTCAGCTNIIIGACC
5′ GGTACGACGTTCAGCTNIIIGACG
5′ GGTACGACGTTCAGCTNIIIGACT

5′ GGTACGACGTTCAGCTNIIIGAGA
5′ GGTACGACGTTCAGCTNIIIGAGC
5′ GGTACGACGTTCAGCTNIIIGAGG
5′ GGTACGACGTTCAGCTNIIIGAGT
5′ GGTACGACGTTCAGCTNIIIGATA
5′ GGTACGACGTTCAGCTNIIIGATC
5′ GGTACGACGTTCAGCTNIIIGATG
5′ GGTACGACGTTCAGCTNIIIGATT

5′ GGTACGACGTTCAGCTNIIIGCAA
5′ GGTACGACGTTCAGCTNIIIGCAC
5′ GGTACGACGTTCAGCTNIIIGCAG
5′ GGTACGACGTTCAGCTNIIIGCAT
5′ GGTACGACGTTCAGCTNIIIGCCA
5′ GGTACGACGTTCAGCTNIIIGCCC
5′ GGTACGACGTTCAGCTNIIIGCCG
5′ GGTACGACGTTCAGCTNIIIGCCT

5′ GGTACGACGTTCAGCTNIIIGCGA
5′ GGTACGACGTTCAGCTNIIIGCGC
5′ GGTACGACGTTCAGCTNIIIGCGG
5′ GGTACGACGTTCAGCTNIIIGCGT
5′ GGTACGACGTTCAGCTNIIIGCTA
5′ GGTACGACGTTCAGCTNIIIGCTC
5′ GGTACGACGTTCAGCTNIIIGCTG
5′ GGTACGACGTTCAGCTNIIIGCTT

5′ GGTACGACGTTCAGCTNIIIGGAA
5′ GGTACGACGTTCAGCTNIIIGGAC
5′ GGTACGACGTTCAGCTNIIIGGAG
5′ GGTACGACGTTCAGCTNIIIGGAT
5′ GGTACGACGTTCAGCTNIIIGGCA
5′ GGTACGACGTTCAGCTNIIIGGCC
5′ GGTACGACGTTCAGCTNIIIGGCG
5′ GGTACGACGTTCAGCTNIIIGGCT

5′ GGTACGACGTTCAGCTNIIIGGGA
5′ GGTACGACGTTCAGCTNIIIGGGC
5′ GGTACGACGTTCAGCTNIIIGGGG
5′ GGTACGACGTTCAGCTNIIIGGGT
5′ GGTACGACGTTCAGCTNIIIGGTA
5′ GGTACGACGTTCAGCTNIIIGCTC
5′ GGTACGACGTTCAGCTNIIIGGTG
5′ GGTACGACGTTCAGCTNIIIGGTT

5′ GGTACGACGTTCAGCTNIIIGTAA

5′ GGTACGACGTTCAGCTNIIIGTAC
5′ GGTACGACGTTCAGCTNIIIGTAG
5′ GGTACGACGTTCAGCTNIIIGTAT
5′ GGTACGACGTTCAGCTNIIIGTCA
5′ GGTACGACGTTCAGCTNIIIGTCC
5′ GGTACGACGTTCAGCTNIIIGTCG
5′ GGTACGACGTTCAGCTNIIIGTCT

5′ GGTACGACGTTCAGCTNIIIGTGA
5′ GGTACGACGTTCAGCTNIIIGTGC
5′ GGTACGACGTTCAGCTNIIIGTGG
5′ GGTACGACGTTCAGCTNIIIGTGT
5′ GGTACGACGTTCAGCTNIIIGTTA
5′ GGTACGACGTTCAGCTNIIIGTTC
5′ GGTACGACGTTCAGCTNIIIGTTG
5′ GGTACGACGTTCAGCTNIIIGTTT

5′ GGTACGACGTTCAGCTNIIITAAA
5′ GGTACGACGTTCAGCTNIIITAAC
5′ GGTACGACGTTCAGCTNIIITAAG
5′ GGTACGACGTTCAGCTNIIITAAT
5′ GGTACGACGTTCAGCTNIIITACA
5′ GGTACGACGTTCAGCTNIIITACC
5′ GGTACGACGTTCAGCTNIIITACG
5′ GGTACGACGTTCAGCTNIIITACT

5′ GGTACGACGTTCAGCTNIIITAGA
5′ GGTACGACGTTCAGCTNIIITAGC
5′ GGTACGACGTTCAGCTNIIITAGG
5′ GGTACGACGTTCAGCTNIIITAGT
5′ GGTACGACGTTCAGCTNIIITATA
5′ GGTACGACGTTCAGCTNIIITATC
5′ GGTACGACGTTCAGCTNIIITATG
5′ GGTACGACGTTCAGCTNIIITATT

5′ GGTACGACGTTCAGCTNIIITCAA
5′ GGTACGACGTTCAGCTNIIITCAC
5′ GGTACGACGTTCAGCTNIIITCAG
5′ GGTACGACGTTCAGCTNIIITCAT
5′ GGTACGACGTTCAGCTNIIITCCA
5′ GGTACGACGTTCAGCTNIIITCCC
5′ GGTACGACGTTCAGCTNIIITCCG
5′ GGTACGACGTTCAGCTNIIITCCT

5′ GGTACGACGTTCAGCTNIIITCGA
5′ GGTACGACGTTCAGCTNIIITCGC
5′ GGTACGACGTTCAGCTNIIITCGG
5′ GGTACGACGTTCAGCTNIIITCGT
5′ GGTACGACGTTCAGCTNIIITCTA
5′ GGTACGACGTTCAGCTNIIITCTC
5′ GGTACGACGTTCAGCTNIIITCTG
5′ GGTACGACGTTCAGCTNIIITCTT

5' GGTACGACGTTCAGCTNIIIAAA
5' GGTACGACGTTCAGCTNIIIAAC
5' GGTACGACGTTCAGCTNIIIAAG
5' GGTACGACGTTCAGCTNIIIAAT
5' GGTACGACGTTCAGCTNIIIACA
5' GGTACGACGTTCAGCTNIIIACC
5' GGTACGACGTTCAGCTNIIIACG
5' GGTACGACGTTCAGCTNIIIACT

5' GGTACGACGTTCAGCTNIIIAGA
5' GGTACGACGTTCAGCTNIIIAGC
5' GGTACGACGTTCAGCTNIIIAGG
5' GGTACGACGTTCAGCTNIIIAGT
5' GGTACGACGTTCAGCTNIIIATA
5' GGTACGACGTTCAGCTNIIIATC
5' GGTACGACGTTCAGCTNIIIATG
5' GGTACGACGTTCAGCTNIIIATT

5' GGTACGACGTTCAGCTNIIICAA
5' GGTACGACGTTCAGCTNIIICAC
5' GGTACGACGTTCAGCTNIIICAG
5' GGTACGACGTTCAGCTNIIICAT
5' GGTACGACGTTCAGCTNIIICCA
5' GGTACGACGTTCAGCTNIIICCC
5' GGTACGACGTTCAGCTNIIICCG
5' GGTACGACGTTCAGCTNIIICCT

5' GGTACGACGTTCAGCTNIIICGA
5' GGTACGACGTTCAGCTNIIICGC
5' GGTACGACGTTCAGCTNIIICGG
5' GGTACGACGTTCAGCTNIIICGT
5' GGTACGACGTTCAGCTNIIICTA
5' GGTACGACGTTCAGCTNIIICTC
5' GGTACGACGTTCAGCTNIIICTG
5' GGTACGACGTTCAGCTNIIICTT

5' GGTACGACGTTCAGCTNIIIGAA
5' GGTACGACGTTCAGCTNIIIGAC
5' GGTACGACGTTCAGCTNIIIGAG
5' GGTACGACGTTCAGCTNIIIGAT
5' GGTACGACGTTCAGCTNIIIGCA
5' GGTACGACGTTCAGCTNIIIGCC
5' GGTACGACGTTCAGCTNIIIGCG
5' GGTACGACGTTCAGCTNIIIGCT

5' GGTACGACGTTCAGCTNIIIGGA
5' GGTACGACGTTCAGCTNIIIGGC
5' GGTACGACGTTCAGCTNIIIGGG
5' GGTACGACGTTCAGCTNIIIGGT

5' GGTACGACGTTCAGCTNIIIGTA
5' GGTACGACGTTCAGCTNIIICTC
5' GGTACGACGTTCAGCTNIIIGTG
5' GGTACGACGTTCAGCTNIIIGTT

5' GGTACGACGTTCAGCTNIIITAA
5' GGTACGACGTTCAGCTNIIITAC
5' GGTACGACGTTCAGCTNIIITAG
5' GGTACGACGTTCAGCTNIIITAT
5' GGTACGACGTTCAGCTNIIITCA
5' GGTACGACGTTCAGCTNIIITCC
5' GGTACGACGTTCAGCTNIIITCG
5' GGTACGACGTTCAGCTNIIITCT

5' GGTACGACGTTCAGCTNIIITGA
5' GGTACGACGTTCAGCTNIIITGC
5' GGTACGACGTTCAGCTNIIITGG
5' GGTACGACGTTCAGCTNIIITGT
5' GGTACGACGTTCAGCTNIIITTA
5' GGTACGACGTTCAGCTNIIITTC
5' GGTACGACGTTCAGCTNIIITTG
5' GGTACGACGTTCAGCTNIIITTT

FIG. 15

5' GGTACGACGTTCAGCTNGGGIIIAAA
5' GGTACGACGTTCAGCTNGGGIIIAAC
5' GGTACGACGTTCAGCTNGGGIIIAAG
5' GGTACGACGTTCAGCTNGGGIIIAAT
5' GGTACGACGTTCAGCTNGGGIIIACA
5' GGTACGACGTTCAGCTNGGGIIIACC
5' GGTACGACGTTCAGCTNGGGIIIACG
5' GGTACGACGTTCAGCTNGGGIIIACT

5' GGTACGACGTTCAGCTNGGGIIIAGA
5' GGTACGACGTTCAGCTNGGGIIIAGC
5' GGTACGACGTTCAGCTNGGGIIIAGG
5' GGTACGACGTTCAGCTNGGGIIIAGT
5' GGTACGACGTTCAGCTNGGGIIIATA
5' GGTACGACGTTCAGCTNGGGIIIATC
5' GGTACGACGTTCAGCTNGGGIIIATG
5' GGTACGACGTTCAGCTNGGGIIIATT

5' GGTACGACGTTCAGCTNGGGIIICAA
5' GGTACGACGTTCAGCTNGGGIIICAC
5' GGTACGACGTTCAGCTNGGGIIICAG
5' GGTACGACGTTCAGCTNGGGIIICAT
5' GGTACGACGTTCAGCTNGGGIIICCA
5' GGTACGACGTTCAGCTNGGGIIICCC
5' GGTACGACGTTCAGCTNGGGIIICCG
5' GGTACGACGTTCAGCTNGGGIIICCT

5' GGTACGACGTTCAGCTNGGGIIICGA
5' GGTACGACGTTCAGCTNGGGIIICGC
5' GGTACGACGTTCAGCTNGGGIIICGG
5' GGTACGACGTTCAGCTNGGGIIICGT
5' GGTACGACGTTCAGCTNGGGIIICTA
5' GGTACGACGTTCAGCTNGGGIIICTC
5' GGTACGACGTTCAGCTNGGGIIICTG
5' GGTACGACGTTCAGCTNGGGIIICTT

5' GGTACGACGTTCAGCTNGGGIIIGAA
5' GGTACGACGTTCAGCTNGGGIIIGAC
5' GGTACGACGTTCAGCTNGGGIIIGAG
5' GGTACGACGTTCAGCTNGGGIIIGAT
5' GGTACGACGTTCAGCTNGGGIIIGCA
5' GGTACGACGTTCAGCTNGGGIIIGCC
5' GGTACGACGTTCAGCTNGGGIIIGCG
5' GGTACGACGTTCAGCTNGGGIIIGCT

5' GGTACGACGTTCAGCTNGGGIIIGGA
5' GGTACGACGTTCAGCTNGGGIIIGGC
5' GGTACGACGTTCAGCTNGGGIIIGGG
5' GGTACGACGTTCAGCTNGGGIIIGGT

5' GGTACGACGTTCAGCTNGGGIIIGTA
5' GGTACGACGTTCAGCTNGGGIIICTC
5' GGTACGACGTTCAGCTNGGGIIIGTG
5' GGTACGACGTTCAGCTNGGGIIIGTT

5' GGTACGACGTTCAGCTNGGGIIITAA
5' GGTACGACGTTCAGCTNGGGIIITAC
5' GGTACGACGTTCAGCTNGGGIIITAG
5' GGTACGACGTTCAGCTNGGGIIITAT
5' GGTACGACGTTCAGCTNGGGIIITCA
5' GGTACGACGTTCAGCTNGGGIIITCC
5' GGTACGACGTTCAGCTNGGGIIITCG
5' GGTACGACGTTCAGCTNGGGIIITCT

5' GGTACGACGTTCAGCTNGGGIIITGA
5' GGTACGACGTTCAGCTNGGGIIITGC
5' GGTACGACGTTCAGCTNGGGIIITGG
5' GGTACGACGTTCAGCTNGGGIIITGT
5' GGTACGACGTTCAGCTNGGGIIITTA
5' GGTACGACGTTCAGCTNGGGIIITTC
5' GGTACGACGTTCAGCTNGGGIIITTG
5' GGTACGACGTTCAGCTNGGGIIITTT

FIG. 16

```
AdaTTTT  5' TTTTGCAGGTACGTCGTACCGCGGCCGCTCGTCGAACGAACACTGCGTT
AdaTTTG  5' TTTGGCAGGTACGTCGTACCGCGGCCGCTACGTCTCGTTCCGCTCTGCA
AdaTTTC  5' TTTCGCAGGTACGTCGTACCGCGGCCGCCTTCGCTAGACGACGTGCGCT
AdaTTTA  5' TTTAGCAGGTACGTCGTACCGCGGCCGCTCGCTCGTACTCCGTCTCAGA
AdaTTGT  5' TTGTGCAGGTACGTCGTACCGCGGCCGCATGCGCGAAAACGGTCACTGT
AdaTTGG  5' TTGGGCAGGTACGTCGTACCGCGGCCGCCGCTTTTCGTCGTCCCAGAGT
AdaTTGC  5' TTGCGCAGGTACGTCGTACCGCGGCCGCCTGGTAAACCCGGTTCGGTCA
AdaTTGA  5' TTGAGCAGGTACGTCGTACCGCGGCCGCCGATAGTCCGAGACGTCTCGT

AdaTTCT  5' TTCTGCAGGTACGTCGTACCGCGGCCGCCTATCGAACGGCACGTTGTTG
AdaTTCG  5' TTCGGCAGGTACGTCGTACCGCGGCCGCCTACGACCGTTCGGGTGGTAC
AdaTTCC  5' TTCCGCAGGTACGTCGTACCGCGGCCGCTCGTACGGTCGACGGTTGTGC
AdaTTCA  5' TTCAGCAGGTACGTCGTACCGCGGCCGCCGTTAGACCTCGTCTCGGCTT
AdaTTAT  5' TTATGCAGGTACGTCGTACCGCGGCCGCCTAACGTCCGATCGCAGCGAT
AdaTTAG  5' TTAGGCAGGTACGTCGTACCGCGGCCGCCGGTATCGGAACGTCTGCGGT
AdaTTAC  5' TTACGCAGGTACGTCGTACCGCGGCCGCTACCGAACCGTCGTACGTAGA
AdaTTAA  5' TTAAGCAGGTACGTCGTACCGCGGCCGCCTAGCGGATTCACGTGGTGTA

AdaTGTT  5' TGTTGCAGGTACGTCGTACCGCGGCCGCGTCGACGACTCTGTCCAGGAT
AdaTGTG  5' TGTGGCAGGTACGTCGTACCGCGGCCGCTCGACGGTTCTTGACTCACGA
AdaTGTC  5' TGTCGCAGGTACGTCGTACCGCGGCCGCCGTAAGGATCACGTGTGCTGT
AdaTGTA  5' TGTAGCAGGTACGTCGTACCGCGGCCGCCTTACGGATCGAGCGTGTACCA
AdaTGGT  5' TGGTGCAGGTACGTCGTACCGCGGCCGCCATACGCAACGACCCGAGTGAG
AdaTGGG  5' TGGGGCAGGTACGTCGTACCGCGGCCGCCGTATCCTTGCGTGTGTAATCC
AdaTGGC  5' TGGCGCAGGTACGTCGTACCGCGGCCGCATCGGAAGAAGCACGATGCCA
AdaTGGA  5' TGGAGCAGGTACGTCGTACCGCGGCCGCCCGATTTCTCGCGAGGACGTC

AdaTGCT  5' TGCTGCAGGTACGTCGTACCGCGGCCGCCGCATCCCACTGCTGGATACT
AdaTGCG  5' TGCGGCAGGTACGTCGTACCGCGGCCGCTACGGTTCGGTCGCTCTTCGT
AdaTGCC  5' TGCCGCAGGTACGTCGTACCGCGGCCGCCATTCGTCGAACCGCAGTTGC
AdaTGCA  5' TGCAGCAGGTACGTCGTACCGCGGCCGCCGAATTTCGACTCGCGAACGT
AdaTGAT  5' TGATGCAGGTACGTCGTACCGCGGCCGCAACGCGACGCAGTGACACGTT
AdaTGAG  5' TGAGGCAGGTACGTCGTACCGCGGCCGCGCTTTGCTTCTCCGTCTAGAC
AdaTGAC  5' TGACGCAGGTACGTCGTACCGCGGCCGCACGTAACGTCGAGCGCTAACG
AdaTGAA  5' TGAAGCAGGTACGTCGTACCGCGGCCGCTACGTGACGTGCGGAGTGATC

AdaTCTT  5' TCTTGCAGGTACGTCGTACCGCGGCCGCCGTAGACGCTCGGTAACTAGT
AdaTCTG  5' TCTGGCAGGTACGTCGTACCGCGGCCGCCTACGAGCGTCTCTCGCATAG
AdaTCTC  5' TCTCGCAGGTACGTCGTACCGCGGCCGCGCGAAACGCATCCGTTCGTGT
AdaTCTA  5' TCTAGCAGGTACGTCGTACCGCGGCCGCTTCGCTTCGTTGCTGCTAGTA
AdaTCGT  5' TCGTGCAGGTACGTCGTACCGCGGCCGCACGATCGACACCGGAGTCTAG
AdaTCGG  5' TCGGGCAGGTACGTCGTACCGCGGCCGCATCGTTGTCGGGACGTAGTCG
AdaTCGC  5' TCGCGCAGGTACGTCGTACCGCGGCCGCACGACCATCGTCTTCCGAGTT
AdaTCGA  5' TCGAGCAGGTACGTCGTACCGCGGCCGCGTCGTCGATGCGTCCATACTT

AdaTCCT  5' TCCTGCAGGTACGTCGTACCGCGGCCGCAACCGAACGTGTCGACCTTCA
AdaTCCG  5' TCCGGCAGGTACGTCGTACCGCGGCCGCTGGTTACGTTTGCGCGGTATT
AdaTCCC  5' TCCCGCAGGTACGTCGTACCGCGGCCGCCGAACCCTGGTCCTGTGTCTA
AdaTCCA  5' TCCAGCAGGTACGTCGTACCGCGGCCGCGTTCGCTCGTCACGAAGTACT
AdaTCAT  5' TCATGCAGGTACGTCGTACCGCGGCCGCCGCAAACCGCAGCTTCTAGTT
AdaTCAG  5' TCAGGCAGGTACGTCGTACCGCGGCCGCTTGCGACCGTTCGCAGCTAGT

AdaTCAC  5' TCACGCAGGTACGTCGTACCGCGGCCGCAGTCGCAGGTAAGCGTATGCT
AdaTCAA  5' TCAAGCAGGTACGTCGTACCGCGGCCGCCGACTATCGTAGGCTCCACGT

AdaTATT  5' TATTGCAGGTACGTCGTACCGCGGCCGCATCGAGCGATTCGAGGTTAGG
AdaTATG  5' TATGGCAGGTACGTCGTACCGCGGCCGCTCGATACCGTAGCGCCGTGGT
AdaTATC  5' TATCGCAGGTACGTCGTACCGCGGCCGCACCGAACTCTCTCGACGGCAT
AdaTATA  5' TATAGCAGGTACGTCGTACCGCGGCCGCTCGATCGTCACTCGTGGTTGA
AdaTAGT  5' TAGTGCAGGTACGTCGTACCGCGGCCGCGCGACTGACGACGAGGATAGT
AdaTAGG  5' TAGGGCAGGTACGTCGTACCGCGGCCGCGTCGTGGCAGACGGATAACCT
AdaTAGC  5' TAGCGCAGGTACGTCGTACCGCGGCCGCATGCGGTTCGTCAGTCGTGGA
AdaTAGA  5' TAGAGCAGGTACGTCGTACCGCGGCCGCCGTATCGATTCCGTCTAGCTC
```

FIG. 18

AdaTACT 5' TACTGCAGGTACGTCGTACCGCGGCCGCACGAAGATCGACGGCTCCGTT
AdaTACG 5' TACGGCAGGTACGTCGTACCGCGGCCGCTCGAACGAAGACCACGTACTG
AdaTACC 5' TACCGCAGGTACGTCGTACCGCGGCCGCTGCTGACTTCGGCGGTGATCG
AdaTACA 5' TACAGCAGGTACGTCGTACCGCGGCCGCGTTCGTAGACGCGAGTCTGAC
AdaTAAT 5' TAATGCAGGTACGTCGTACCGCGGCCGCCAATCGCGTTTACTCGAATGG
AdaTAAG 5' TAAGGCAGGTACGTCGTACCGCGGCCGCCGATTCGGAACGAGCGACTAT
AdaTAAC 5' TAACGCAGGTACGTCGTACCGCGGCCGCCGTTCTTCCTGGCTCTAAGGC
AdaTAAA 5' TAAAGCAGGTACGTCGTACCGCGGCCGCGAACGCTAACGTGCGTGAACC

AdaGTTT 5' GTTTGCAGGTACGTCGTACCGCGGCCGCGATGAGCCAGGGAAGGAATTGG
AdaGTTG 5' GTTGGCAGGTACGTCGTACCGCGGCCGCGATTAGACGAGGCGACCGAGA
AdaGTTC 5' GTTCGCAGGTACGTCGTACCGCGGCCGCGACTAGGTCAGCTGAGGAGTCT
AdaGTTA 5' GTTAGCAGGTACGTCGTACCGCGGCCGCACTACGAGTGGCGAGCTGACT
AdaGTGT 5' GTGTGCAGGTACGTCGTACCGCGGCCGCGCCTACTGACTCGAGTACTCT
AdaGTGG 5' GTGGGCAGGTACGTCGTACCGCGGCCGCCGTTGTTCCCGACTCCTCTGT
AdaGTGC 5' GTGCGCAGGTACGTCGTACCGCGGCCGCCAACGGACACAGGTACCCAAG
AdaGTGA 5' GTGAGCAGGTACGTCGTACCGCGGCCGCTCACAGATTGCTTCGTACTGT

AdaGTCT 5' GTCTGCAGGTACGTCGTACCGCGGCCGCCGTCCTTATGCTCTCCGAGCT
AdaGTCG 5' GTCGGCAGGTACGTCGTACCGCGGCCGCCCTATCCAGTAATGCCCGAGA
AdaGTCC 5' GTCCGCAGGTACGTCGTACCGCGGCCGCCATGGGACAGATCAGGTTAAGG
AdaGTCA 5' GTCAGCAGGTACGTCGTACCGCGGCCGCTCAGTCAAACTGGACTCGCAC
AdaGTAT 5' GTATGCAGGTACGTCGTACCGCGGCCGCGAAGGCTGTGGTTGGCCACC
AdaGTAG 5' GTAGGCAGGTACGTCGTACCGCGGCCGCGCAGTCGTACACCTCATCTCA
AdaGTAC 5' GTACGCAGGTACGTCGTACCGCGGCCGCTAGACAACTGGCGCATTGACA
AdaGTAA 5' GTAAGCAGGTACGTCGTACCGCGGCCGCGATGCCACTAGGTGTTGGTGC

AdaGGTT 5' GGTTGCAGGTACGTCGTACCGCGGCCGCAGACCTCCCGCTCTACAGATC
AdaGGTG 5' GGTGGCAGGTACGTCGTACCGCGGCCGCGCGTGGCAGCACCCTTAGCCA
AdaGGTC 5' GGTCGCAGGTACGTCGTACCGCGGCCGCAGGTTTCGGCTGGAGTCTCAT
AdaGGTA 5' GGTAGCAGGTACGTCGTACCGCGGCCGCTCTGCAGGTCGTCCCTACACT
AdaGGGT 5' GGGTGCAGGTACGTCGTACCGCGGCCGCTGTATCAGCTATGCCGTTGCC
AdaGGGG 5' GGGGGCAGGTACGTCGTACCGCGGCCGCCGTTGCCTCATACCTCAGCATC
AdaGGGC 5' GGGCGCAGGTACGTCGTACCGCGGCCGCAGCAGACAGGCCAAATGACGT
AdaGGGA 5' GGGAGCAGGTACGTCGTACCGCGGCCGCAGCAGTTCGAGCACGTGAGTC

AdaGGCT 5' GGCTGCAGGTACGTCGTACCGCGGCCGCGAGATGTGCCACATGCCGACT
AdaGGCG 5' GGCGGCAGGTACGTCGTACCGCGGCCGCGTTAGCGGCTGTCTACCGAGT
AdaGGCC 5' GGCCGCAGGTACGTCGTACCGCGGCCGCGTGAATAGTGACCGTGTCCCAG
AdaGGCA 5' GGCAGCAGGTACGTCGTACCGCGGCCGCTGTGTACAGCGTTCCAAGGGT
AdaGGAT 5' GGATGCAGGTACGTCGTACCGCGGCCGCACGTCTCACTACGTGTCCGGA
AdaGGAG 5' GGAGGCAGGTACGTCGTACCGCGGCCGCGCTAGGTAAGAGTGCAAGGCA
AdaGGAC 5' GGACGCAGGTACGTCGTACCGCGGCCGCCCAAGTGCACGATACTGGCTA
AdaGGAA 5' GGAAGCAGGTACGTCGTACCGCGGCCGCTTCGTGTAGGGTGGTAAGCTT

AdaGCTT 5' GCTTGCAGGTACGTCGTACCGCGGCCGCTTCGACCCATGTCATCATCCT
AdaGCTG 5' GCTGGCAGGTACGTCGTACCGCGGCCGCGGAGGACTTGTCATGGAGGTT
AdaGCTC 5' GCTCGCAGGTACGTCGTACCGCGGCCGCAGATGTCTGTCCCACGGTGTT
AdaGCTA 5' GCTAGCAGGTACGTCGTACCGCGGCCGCAATGCCCTGTATGTTCCGTGG
AdaGCGT 5' GCGTGCAGGTACGTCGTACCGCGGCCGCATGCTTAGCTCCACGCTCATG
AdaGCGG 5' GCGGGCAGGTACGTCGTACCGCGGCCGCCCTCGTACCTTGTGACAGATG
AdaGCGC 5' GCGCGCAGGTACGTCGTACCGCGGCCGCTGTGCGAGCGTTCCATCTCCT
AdaGCGA 5' GCGAGCAGGTACGTCGTACCGCGGCCGCGATGCTGGTCGAGGACACAGA

AdaGCCT 5' GCCTGCAGGTACGTCGTACCGCGGCCGCAACCAGCTTCCACCTGGAGTT
AdaGCCG 5' GCCGGCAGGTACGTCGTACCGCGGCCGCGTGACGTTCGATGTGAGTCTG
AdaGCCC 5' GCCCGCAGGTACGTCGTACCGCGGCCGCGGTGTCCCATTCGGAGTTACA
AdaGCCA 5' GCCAGCAGGTACGTCGTACCGCGGCCGCCAGGTGAGGGAGAGCATCAAC
AdaGCAT 5' GCATGCAGGTACGTCGTACCGCGGCCGCCTCGACCTTGTTGGCAGACTC
AdaGCAG 5' GCAGGCAGGTACGTCGTACCGCGGCCGCTGCATTGCGGTTCCTCACAAC
AdaGCAC 5' GCACGCAGGTACGTCGTACCGCGGCCGCCTGAGAGCTGTTCACTGAGGT
AdaGCAA 5' GCAAGCAGGTACGTCGTACCGCGGCCGCCAGTGGATAGCGACCTGGATC

AdaGATT 5' GATTGCAGGTACGTCGTACCGCGGCCGCTGGATTTGAGATCGCATGTGG
AdaGATG 5' GATGGCAGGTACGTCGTACCGCGGCCGCGCGTCTAGGTCGTCACGACAA
AdaGATC 5' GATCGCAGGTACGTCGTACCGCGGCCGCCCCGATGATTAGACGCCTCAA
AdaGATA 5' GATAGCAGGTACGTCGTACCGCGGCCGCGATAGTGGCGATTGCTGAGCG
AdaGAGT 5' GAGTGCAGGTACGTCGTACCGCGGCCGCGGCGGCTCATGTGTGATAATG
AdaGAGG 5' GAGGGCAGGTACGTCGTACCGCGGCCGCGGTGGTTTTCTCTTGCCCCTC
AdaGAGC 5' GAGCGCAGGTACGTCGTACCGCGGCCGCGGAAGACAGCTCGCTGGACTT
AdaGAGA 5' GAGAGCAGGTACGTCGTACCGCGGCCGCGACCTTACTGAGAAGCGGCCT

AdaGACT 5' GACTGCAGGTACGTCGTACCGCGGCCGCTGTGGGGTTCTCTGCGTACAC
AdaGACG 5' GACGGCAGGTACGTCGTACCGCGGCCGCGGAACTGTTGTCCAGGAGACC
AdaGACC 5' GACCGCAGGTACGTCGTACCGCGGCCGCGATGCCAGAGTGATGCAACTG
AdaGACA 5' GACAGCAGGTACGTCGTACCGCGGCCGCCTCATTTCGCGATTCGGTCTT
AdaGAAT 5' GAATGCAGGTACGTCGTACCGCGGCCGCTGTGAACCGGGTGGTCTGAAC
AdaGAAG 5' GAAGGCAGGTACGTCGTACCGCGGCCGCCCAGAGGATCCTGCTGTAGCA
AdaGAAC 5' GAACGCAGGTACGTCGTACCGCGGCCGCGTTGCTGCCGGTCTAGGTATC
AdaGAAA 5' GAAAGCAGGTACGTCGTACCGCGGCCGCGTGGGTTTCCTAGCTCTCAGTG

AdaCTTT 5' CTTTGCAGGTACGTCGTACCGCGGCCGCGACGGCATCAGGATTTATCCC
AdaCTTg 5' CTTgGCAGGTACGTCGTACCGCGGCCGCGACAGGCAAGGGTTGAGACTG
AdaCTTc 5' CTTcGCAGGTACGTCGTACCGCGGCCGCGTGCTTAGGTGAGCTCCGTGA
AdaCTTa 5' CTTaGCAGGTACGTCGTACCGCGGCCGCGCATCAGGTCAGCTCCAAATC
AdaCTGT 5' CTGTGCAGGTACGTCGTACCGCGGCCGCACTCGATTTCATAGATACATC
AdaCTGg 5' CTGgGCAGGTACGTCGTACCGCGGCCGCCTGAACGGAGTTACGGGGTTG
AdaCTGc 5' CTGcGCAGGTACGTCGTACCGCGGCCGCCGAAGGTCGCACAAGTCGACG
AdaCTGa 5' CTGaGCAGGTACGTCGTACCGCGGCCGCCATGACCTGGTTGTGCGTGTT

AdaCTCT 5' CTCTGCAGGTACGTCGTACCGCGGCCGCGAGTCCAGTAGCTGGAGACGA
AdaCTCg 5' CTCgGCAGGTACGTCGTACCGCGGCCGCCGTCTGTAAGTGTTGAGCGTA
AdaCTCc 5' CTCcGCAGGTACGTCGTACCGCGGCCGCGTATAGTAGGTGCTGTTGCCC
AdaCTCa 5' CTCaGCAGGTACGTCGTACCGCGGCCGCGAGTGGAAGACTGGTCAGACG
AdaCTAT 5' CTATGCAGGTACGTCGTACCGCGGCCGCCGACTGGAAAACACCTCTCCC
AdaCTAg 5' CTAgGCAGGTACGTCGTACCGCGGCCGCTTCGCGACGACTACTTGGGTC
AdaCTAc 5' CTAcGCAGGTACGTCGTACCGCGGCCGCGGAGCCTGATTCACCATGGTG
AdaCTAa 5' CTAaGCAGGTACGTCGTACCGCGGCCGCTCGACCTTCCCACACTCAGGT

AdaCGTT 5' CGTTGCAGGTACGTCGTACCGCGGCCGCCTTCCGACAGACAGTTTCGCA
AdaCGTg 5' CGTgGCAGGTACGTCGTACCGCGGCCGCCTGCTTGAGGATTGCCAAAGC
AdaCGTc 5' CGTcGCAGGTACGTCGTACCGCGGCCGCTGTTCGATGAGTCAGGAGCGA
AdaCGTa 5' CGTaGCAGGTACGTCGTACCGCGGCCGCGTCGAGATGGTACCGACTCTG
AdaCGGT 5' CGGTGCAGGTACGTCGTACCGCGGCCGCGATCAACATATCCTTTGTCCGC
AdaCGGg 5' CGGgGCAGGTACGTCGTACCGCGGCCGCCTTCCGACAAAAGGTCCTAGTG
AdaCGGc 5' CGGcGCAGGTACGTCGTACCGCGGCCGCCTCACGCTAGCACCGCTCTTA
AdaCGGa 5' CGGaGCAGGTACGTCGTACCGCGGCCGCACCAACGAATCTCGAGGCTCC

AdaCGCT 5' CGCTGCAGGTACGTCGTACCGCGGCCGCGTCCATGCTGAAAGACCAGGC
AdaCGCg 5' CGCgGCAGGTACGTCGTACCGCGGCCGCGACAGACGTTCGGCTGACAAG
AdaCGCc 5' CGCcGCAGGTACGTCGTACCGCGGCCGCCTGTTATCCGCCTTGCAGCGT
AdaCGCa 5' CGCaGCAGGTACGTCGTACCGCGGCCGCTGAGGCACCTGTCTATGTGCTG
AdaCGAT 5' CGATGCAGGTACGTCGTACCGCGGCCGCCCCTTGGTGCAACGTGATGAT
AdaCGAg 5' CGAgGCAGGTACGTCGTACCGCGGCCGCAGGGTCGTACGTTCCTCTGGA
AdaCGAc 5' CGAcGCAGGTACGTCGTACCGCGGCCGCTGCTCGCTCTGTCTCCTGTTC
AdaCGAa 5' CGAaGCAGGTACGTCGTACCGCGGCCGCTCCACGTCGTACTGACCGTAG

AdaCCTT 5' CCTTGCAGGTACGTCGTACCGCGGCCGCCAGCGCTCCAGCGAATAATTT
AdaCCTg 5' CCTgGCAGGTACGTCGTACCGCGGCCGCCTGTATGCTACACCGCAGCTG
AdaCCTc 5' CCTcGCAGGTACGTCGTACCGCGGCCGCGCGTCTGGAACGATTACGAGT
AdaCCTa 5' CCTaGCAGGTACGTCGTACCGCGGCCGCCTTTCCGGCCATAACACGTAT
AdaCcgT 5' CcgTGCAGGTACGTCGTACCGCGGCCGCTTCCCAACCTAACTTCGGAGC
AdaCcgg 5' CcggGCAGGTACGTCGTACCGCGGCCGCGCTATCGTTGCAAGGGTGAAC
AdaCcgc 5' CcgcGCAGGTACGTCGTACCGCGGCCGCATCCGAGTAGGGAGTCTCCCT
AdaCcga 5' CcgaGCAGGTACGTCGTACCGCGGCCGCCTGGTACATGTCTGGCAACTC

AdaCcCt 5' CcCtGCAGGTACGTCGTACCGCGGCCGCTGCTTGTGGGAAGAAACTGGC
AdaCcCg 5' CcCgGCAGGTACGTCGTACCGCGGCCGCCTGGGTCGCATACTCGTTCAC
AdaCcCc 5' CcCcGCAGGTACGTCGTACCGCGGCCGCTAGGCGTTGCAATACAGGCAA
AdaCcCa 5' CcCaGCAGGTACGTCGTACCGCGGCCGCTCCTGATCTTCCAGCAGCTCA
AdaCcAt 5' CcAtGCAGGTACGTCGTACCGCGGCCGCTCATGTTGCGTATTCCCCTACT
AdaCcAg 5' CcAgGCAGGTACGTCGTACCGCGGCCGCGGTCACTCCTCGAAGTGGATC
AdaCcAc 5' CcAcGCAGGTACGTCGTACCGCGGCCGCCGTGTCGAGCGAATGTGGTAC
AdaCcAa 5' CcAaGCAGGTACGTCGTACCGCGGCCGCTGCCAATTCGAGAGGGAAATG

AdaCaTT 5' CaTTGCAGGTACGTCGTACCGCGGCCGCACATCTCGGCGATACATCGGA
AdaCaTg 5' CaTgGCAGGTACGTCGTACCGCGGCCGCTGACCTTCCTTCTCAGTTGGT
AdaCaTc 5' CaTcGCAGGTACGTCGTACCGCGGCCGCGAGTACTGGTGACTAGGATTCG
AdaCaTa 5' CaTaGCAGGTACGTCGTACCGCGGCCGCTCGGTAGGAGTGCATCCTTCC
AdaCagT 5' CagTGCAGGTACGTCGTACCGCGGCCGCGTGGTGAGGGAGGGAGCTATT
AdaCagg 5' CaggGCAGGTACGTCGTACCGCGGCCGCGAGATGTAGCAGATGGTCCTC
AdaCagc 5' CagcGCAGGTACGTCGTACCGCGGCCGCGATTAGGACCCGCGAAGTAAT
AdaCaga 5' CagaGCAGGTACGTCGTACCGCGGCCGCCATAGCCTTTTGTTGTCGCTG

AdaCaCT 5' CaCTGCAGGTACGTCGTACCGCGGCCGCCTAGTGATGAGGTGTCAACGC
AdaCaCg 5' CaCgGCAGGTACGTCGTACCGCGGCCGCGTTCGGACGGTGAGTGGTAGA
AdaCaCc 5' CaCcGCAGGTACGTCGTACCGCGGCCGCGCACTTCCTCAGAAGCATGCT
AdaCaCa 5' CaCaGCAGGTACGTCGTACCGCGGCCGCTGTAACACAAGCCACGATGACT
AdaCaAT 5' CaATGCAGGTACGTCGTACCGCGGCCGCTGGGACCCATAGCCAAGTGTC
AdaCaAg 5' CaAgGCAGGTACGTCGTACCGCGGCCGCTTGGAGCTCTCACGTACTGTG
AdaCaAc 5' CaAcGCAGGTACGTCGTACCGCGGCCGCCTGGTCCTGTTGGGTTGCTTC
AdaCaAa 5' CaAaGCAGGTACGTCGTACCGCGGCCGCTGCTGCTTCCTGGTGAGTCTCT

AdaATTT 5' ATTTGCAGGTACGTCGTACCGCGGCCGCTCAGCCTTCTGCTGTACCACA
AdaATTG 5' ATTGGCAGGTACGTCGTACCGCGGCCGCTCCTGCTTTGGCGAACTTGGA
AdaATTC 5' ATTCGCAGGTACGTCGTACCGCGGCCGCGAAGCATTCCAGAATCGCACG
AdaATTA 5' ATTAGCAGGTACGTCGTACCGCGGCCGCTCCTCTCCCACCATTCCGAGT
AdaATGT 5' ATGTGCAGGTACGTCGTACCGCGGCCGCTGGTTTCCTCGAGAATCTGCT
AdaATGG 5' ATGGGCAGGTACGTCGTACCGCGGCCGCTGCTGACCAGACCAGAGAGGT
AdaATGC 5' ATGCGCAGGTACGTCGTACCGCGGCCGCCTGCTCACTCATCCACTTGTCA
AdaATGA 5' ATGAGCAGGTACGTCGTACCGCGGCCGCTGTACCACGGTCGAAAAATCC

AdaATCT 5' ATCTGCAGGTACGTCGTACCGCGGCCGCTGGTGAGTAAAGCTCTCAGGC
AdaATCG 5' ATCGGCAGGTACGTCGTACCGCGGCCGCGTGCCATCGCAGAAGCTAAGC
AdaATCC 5' ATCCGCAGGTACGTCGTACCGCGGCCGCTGAGGCTACCTGACACTGCTG
AdaATCA 5' ATCAGCAGGTACGTCGTACCGCGGCCGCTGGTGCCAGAACCTCTTGTCT
AdaATAT 5' ATATGCAGGTACGTCGTACCGCGGCCGCTGCGACTTGATAGACCCACCA
AdaATAG 5' ATAGGCAGGTACGTCGTACCGCGGCCGCGTAGCATCAGTGGCTACGACA
AdaATAC 5' ATACGCAGGTACGTCGTACCGCGGCCGCGAGCATCGATGCCAGGGATGA
AdaATAA 5' ATAAGCAGGTACGTCGTACCGCGGCCGCCTCTGTCTGGGAAGGCACCTC

AdaAGTT 5' AGTTGCAGGTACGTCGTACCGCGGCCGCGAGAGGACGACTCCAGAGACG
AdaAGTG 5' AGTGGCAGGTACGTCGTACCGCGGCCGCCATCCATAGACCGTGACTCCA
AdaAGTC 5' AGTCGCAGGTACGTCGTACCGCGGCCGCGTCCACGTCATTCCAGGAGACT
AdaAGTA 5' AGTAGCAGGTACGTCGTACCGCGGCCGCCGATCTAGGTGAGGACACTGG
AdaAGGT 5' AGGTGCAGGTACGTCGTACCGCGGCCGCGAAGGACAACTGGCGTAGGCT
AdaAGGG 5' AGGGGCAGGTACGTCGTACCGCGGCCGCGTCGTTCGTATCCTCTACAAGG
AdaAGGC 5' AGGCGCAGGTACGTCGTACCGCGGCCGCGAGTTCCTGGCTGTGCTACCT
AdaAGGA 5' AGGAGCAGGTACGTCGTACCGCGGCCGCCTGCTGATGGATCGTGTACGA

AdaAGCT 5' AGCTGCAGGTACGTCGTACCGCGGCCGCGTGTTCGAAGGCAGAGGGTTC
AdaAGCG 5' AGCGGCAGGTACGTCGTACCGCGGCCGCGGTTTAAGGGGTCTAGGTTGAG
AdaAGCC 5' AGCCGCAGGTACGTCGTACCGCGGCCGCCTCCTCTGGGACGTTGCTAAG
AdaAGCA 5' AGCAGCAGGTACGTCGTACCGCGGCCGCCAGCAGACGCAGTGATCTCCT
AdaAGAT 5' AGATGCAGGTACGTCGTACCGCGGCCGCGAGCATGAAGTGAGCAGCTGG
AdaAGAG 5' AGAGGCAGGTACGTCGTACCGCGGCCGCCTGGCCTCAGTAGAGACTGGT
AdaAGAC 5' AGACGCAGGTACGTCGTACCGCGGCCGCTGGCAGGCACTAGTGCAATAG
AdaAGAA 5' AGAAGCAGGTACGTCGTACCGCGGCCGCCTACATCCTGTCGTGTCCGCA

AdaACTT 5' ACTTGCAGGTACGTCGTACCGCGGCCGCACGGATCTGAGGTAGCACTGG
AdaACTG 5' ACTGGCAGGTACGTCGTACCGCGGCCGCTGAGAAGCCTCAGCTCGATTC
AdaACTC 5' ACTCGCAGGTACGTCGTACCGCGGCCGCCGTTTGAGGTTAATTCGCCTG
AdaACTA 5' ACTAGCAGGTACGTCGTACCGCGGCCGCTGCTACTGCAAATACCCGAGC
AdaACGT 5' ACGTGCAGGTACGTCGTACCGCGGCCGCTCGTCTGCACGTAACTCCTCA
AdaACGG 5' ACGGGCAGGTACGTCGTACCGCGGCCGCGGATGGTTCAAGCGACTGTCA
AdaACGC 5' ACGCGCAGGTACGTCGTACCGCGGCCGCCACTCTCAGCTGGTGGTCTGT
AdaACGA 5' ACGAGCAGGTACGTCGTACCGCGGCCGCGCTTGGTTGATGGGAATGGAC

AdaACCT 5' ACCTGCAGGTACGTCGTACCGCGGCCGCCTGCAAGCAAGACTAGACGTAC
AdaACCG 5' ACCGGCAGGTACGTCGTACCGCGGCCGCTCCATGAAGCTCCCAAGTGTC
AdaACCC 5' ACCCGCAGGTACGTCGTACCGCGGCCGCTGCACATTGGTTAAACGCAGG
AdaACCA 5' ACCAGCAGGTACGTCGTACCGCGGCCGCTGCAGATCATTGCAGGTAGAT
AdaACAT 5' ACATGCAGGTACGTCGTACCGCGGCCGCAGCACGACTGGAAGACGGTCT
AdaACAG 5' ACAGGCAGGTACGTCGTACCGCGGCCGCAGCATGAACATCATCGGTGGT
AdaACAC 5' ACACGCAGGTACGTCGTACCGCGGCCGCTGGAGTCTGGACTGTGGTGGA
AdaACAA 5' ACAAGCAGGTACGTCGTACCGCGGCCGCCTTACACCTCGTCGACTCGTC

AdaAATT 5' AATTGCAGGTACGTCGTACCGCGGCCGCAGGGCTAAGCTTCCGTAGGTC
AdaAATG 5' AATGGCAGGTACGTCGTACCGCGGCCGCGTTGGCTAGTTGCGGTGGTGT
AdaAATC 5' AATCGCAGGTACGTCGTACCGCGGCCGCTCGCAGGTCTTAGGCACAACG
AdaAATA 5' AATAGCAGGTACGTCGTACCGCGGCCGCGTCAGGTCCGATTCTGGTTCC
AdaAAGT 5' AAGTGCAGGTACGTCGTACCGCGGCCGCCAGCTCGACCCATCAACTCCA
AdaAAGG 5' AAGGGCAGGTACGTCGTACCGCGGCCGCGCTACCGGCAACATAGCTGTC
AdaAAGC 5' AAGCGCAGGTACGTCGTACCGCGGCCGCTCCTCAGTATGTGCCACTCTGA
AdaAAGA 5' AAGAGCAGGTACGTCGTACCGCGGCCGCGTTGCTTGCCTACGTGCCATC

AdaAACT 5' AACTGCAGGTACGTCGTACCGCGGCCGCGACTAGGCCCGAAGCACAGAG
AdaAACG 5' AACGGCAGGTACGTCGTACCGCGGCCGCTTGCTGGACACAGGTGATACG
AdaAACC 5' AACCGCAGGTACGTCGTACCGCGGCCGCTGGAACGCCACCGTTGTTAG
AdaAACA 5' AACAGCAGGTACGTCGTACCGCGGCCGCCGTATGGGATCGAGGTTGCA
AdaAAAT 5' AAATGCAGGTACGTCGTACCGCGGCCGCTCAGTCTCATCAGCTCCTCAC
AdaAAAG 5' AAAGGCAGGTACGTCGTACCGCGGCCGCCCTTGGTGACGTAACCGTTCG
AdaAAAC 5' AAACGCAGGTACGTCGTACCGCGGCCGCGACGAGCCTAGGAACTGAAGT
AdaAAAA 5' AAAAGCAGGTACGTCGTACCGCGGCCGCGTGGAGTCGTGATGGAGACCT

```
                |------constant-----|------specific---------|
AdaTTT  5'-Ph-TTTNAGCTGAACGTCGTACC CGTCG AACGA ACACG GGCGT
AdaTTG  5'-Ph-TTGNAGCTGAACGTCGTACC TACGC TCGTT CCGCT CTGCG
AdaTTC  5'-Ph-TTCNAGCTGAACGTCGTACC CGCGA GACGA CGTGC GCGGC
AdaTTA  5'-Ph-TTANAGCTGAACGTCGTACC TCGCG TCGTC CCCGT CGCAG
AdaTGT  5'-Ph-TGTNAGCTGAACGTCGTACC ACGCG CGAAA ACGGG CACCG
AdaTGG  5'-Ph-TGGNAGCTGAACGTCGTACC CGCGT TTTCG CCGTC CGAGA
AdaTGC  5'-Ph-TGCNAGCTGAACGTCGTACC CGTAC CGGAC GCCGA CGCCA
AdaTGA  5'-Ph-TGANAGCTGAACGTCGTACC CGATA GTCCG AGACG TCTCG

AdaTCT  5'-Ph-TCTNAGCTGAACGTCGTACC TATCG AACGG CACGT TGGCG
AdaTCG  5'-Ph-TCGNAGCTGAACGTCGTACC TACGA CCGTT CGGGT GGTAC
AdaTCC  5'-Ph-TCCNAGCTGAACGTCGTACC TCGTA CGGTC GACGG CCGTG
AdaTCA  5'-Ph-TCANAGCTGAACGTCGTACC CGTTA GACCG CGTCT CGGCT
AdaTAT  5'-Ph-TATNAGCTGAACGTCGTACC TAACG TCCGA TCGGG CAGCG
AdaTAG  5'-Ph-TAGNAGCTGAACGTCGTACC CGGTA TCGGA ACGCC GCCGC
AdaTAC  5'-Ph-TACNAGCTGAACGTCGTACC TACCG ATCCG TCGGC CGGAG
AdaTAA  5'-Ph-TAANAGCTGAACGTCGTACC TAGCG CGGAT TCACG CGGTG

AdaGTT  5'-Ph-GTTNAGCTGAACGTCGTACC GTCGA CGACC CCCGA CACGG
AdaGTG  5'-Ph-GTGNAGCTGAACGTCGTACC TCGAC GGTCG CGTGA CTCCG
AdaGTC  5'-Ph-GTCNAGCTGAACGTCGTACC CGTAA CGATC ACGTG CGCTG
AdaGTA  5'-Ph-GTANAGCTGAACGTCGTACC TTACG GATCG AGCGG GTACC
AdaGGT  5'-Ph-GGTNAGCTGAACGTCGTACC ATACG CAACG ACCCG AGCCG
AdaGGG  5'-Ph-GGGNAGCTGAACGTCGTACC CGTAT CGTTG CGTGT GTAAC
AdaGGC  5'-Ph-GGCNAGCTGAACGTCGTACC ATCGG CCGAA GCACG GCGCC
AdaGGA  5'-Ph-GGANAGCTGAACGTCGTACC CCGAT TTCGG GCGAG GGCGC

AdaGCT  5'-Ph-GCTNAGCTGAACGTCGTACC CCGTA CCGAC TGCGG GATAC
AdaGCG  5'-Ph-GCGNAGCTGAACGTCGTACC TACGG GTCGG TCGCT CCTCG
AdaGCC  5'-Ph-GCCNAGCTGAACGTCGTACC ATTCG TCGAA CCGCA GTTAC
AdaGCA  5'-Ph-GCANAGCTGAACGTCGTACC CGAAT TTCGA CTCGC GCCCG
AdaGAT  5'-Ph-GATNAGCTGAACGTCGTACC AACGC GACGC AGCGA CACGC
AdaGAG  5'-Ph-GAGNAGCTGAACGTCGTACC GCGTT GCGTC TCCGC TAGAC
AdaGAC  5'-Ph-GACNAGCTGAACGTCGTACC ACGTA ACGTC GAGCG CTAAC
AdaGAA  5'-Ph-GAANAGCTGAACGTCGTACC TACGT GACGT GCGGA GTATC

AdaCTT  5'-Ph-CTTNAGCTGAACGTCGTACC CGTAG ACGCT CGGCA ACTAG
AdaCTG  5'-Ph-CTGNAGCTGAACGTCGTACC CTACG AGCGT CGCTC GCATA
AdaCTC  5'-Ph-CTCNAGCTGAACGTCGTACC GCGAA ACGCA TCCGG GCGTG
AdaCTA  5'-Ph-CTANAGCTGAACGTCGTACC TTCGC TGCGT TGCCG CTAGT
AdaCGT  5'-Ph-CGTNAGCTGAACGTCGTACC ACGAT CGACA CCGGA GTCTA
AdaCGG  5'-Ph-CGGNAGCTGAACGTCGTACC ATCGT TGTCG GGACG TAGTC
AdaCGC  5'-Ph-CGCNAGCTGAACGTCGTACC ACGAC CATCG TCGCC CGAGG
AdaCGA  5'-Ph-CGANAGCTGAACGTCGTACC GTCGT CGATG CGTCC ATACC
```

```
AdaCCT  5'-Ph-CCTNAGCTGAACGTCGTACC AACCG AACGT GGCGA CCGGC
AdaCCG  5'-Ph-CCGNAGCTGAACGTCGTACC CGGTT ACGTT TGCGC GGTAT
AdaCCC  5'-Ph-CCCNAGCTGAACGTCGTACC CGAAC CCGCG TCGTG GCCGG
AdaCCA  5'-Ph-CCANAGCTGAACGTCGTACC GTTCG CGCGG CACGA AGTAC
AdaCAT  5'-Ph-CATNAGCTGAACGTCGTACC CGCAA ACCGG CGCTT CTAGC

AdaCAG  5'-Ph-CAGNAGCTGAACGTCGTACC TTGCG ACCGG GCGCA GCTAG
AdaCAC  5'-Ph-CACNAGCTGAACGTCGTACC AGTCG CCGGT AAGCG TATGC
AdaCAA  5'-Ph-CAANAGCTGAACGTCGTACC CGACT ATCGC GCGCT CCACG

AdaATT  5'-Ph-ATTNAGCTGAACGTCGTACC ATCGA GCGAT TCGAG GTTA
AdaATG  5'-Ph-ATGNAGCTGAACGTCGTACC TCGAT ACCGT AGCGC CGTGG
AdaATC  5'-Ph-ATCNAGCTGAACGTCGTACC ACCGA ACGGT CTCGA CGGCC
AdaATA  5'-Ph-ATANAGCTGAACGTCGTACC TCGGT CGTCA CTCGT GGTTA
AdaAGT  5'-Ph-AGTNAGCTGAACGTCGTACC GCGAC TGACG ACGAG GATAG
AdaAGG  5'-Ph-AGGNAGCTGAACGTCGTACC GTCGC CGCCG ACGGA TAACC
AdaAGC  5'-Ph-AGCNAGCTGAACGTCGTACC ATGCG CGGCG TCCGT CGGGA
AdaAGA  5'-Ph-AGANAGCTGAACGTCGTACC CGCAT CGCGC GCCGT CTATC

AdaACT  5'-Ph-ACTNAGCTGAACGTCGTACC ACGAA GCGCG ACGGC TCCCG
AdaACG  5'-Ph-ACGNAGCTGAACGTCGTACC TCGCA CGAAG ACCGC GTACT
AdaACC  5'-Ph-ACCNAGCTGAACGTCGTACC TGCGA CTTCG GCGGT GGCCG
AdaACA  5'-Ph-ACANAGCTGAACGTCGTACC TTCGT AAACG CGAGT CTAAC
AdaAAT  5'-Ph-AATNAGCTGAACGTCGTACC AATCG CGTTT ACTCG GGCGG
AdaAAG  5'-Ph-AAGNAGCTGAACGTCGTACC CGATT CGGAA CGAGC GACTA
AdaAAC  5'-Ph-AACNAGCTGAACGTCGTACC CGTTC TTCCG GCTCG CGGGC
AdaAAA  5'-Ph-AAANAGCTGAACGTCGTACC GAACG CGCAC GTGCG TAACC
```

Helper Oligonucleotides

```
UniL1   5' GGTACGACGTTCAGCT
UniL2   5' GGTACGACGTTCAGCA
UniL3   5' GGTACGACGTTCAGAT
```

FIG. 19

5′ Ph-TTTNGCAGGTACGTCGTACC GCGGCCGC GTGAGCTTGAGTCGCGTGGA
5′ Ph-TTGNGCAGGTACGTCGTACC GCGGCCGC CCAACGTCGCGAGTTAGTCAG
5′ Ph-TTCNGCAGGTACGTCGTACC GCGGCCGC AGGTAGACGCGGTATGTTCGTA
5′ Ph-TTANGCAGGTACGTCGTACC GCGGCCGC CGGTGCTAGAGTCGCGTGTT
5′ Ph-TGTNGCAGGTACGTCGTACC GCGGCCGC CGACAGTACCGCGACAGCTA
5′ Ph-TGGNGCAGGTACGTCGTACC GCGGCCGC GCACTTAACTACGCCGACGAAG
5′ Ph-TGCNGCAGGTACGTCGTACC GCGGCCGC gTACTAGCCTAACCGAGGCGTA
5′ Ph-TGANGCAGGTACGTCGTACC GCGGCCGC TCGGATCACGTACACGTGCT

5′ Ph-TCTNGCAGGTACGTCGTACC GCGGCCGC GTACGTCGCCTAGTCGACCTG
5′ Ph-TCGNGCAGGTACGTCGTACC GCGGCCGC cTCTCCTAACGGACCGACTAAC
5′ Ph-TCCNGCAGGTACGTCGTACC GCGGCCGC CGTTCCGATCTAGCGGTATCTT
5′ Ph-TCANGCAGGTACGTCGTACC GCGGCCGC gcACCCGTACaGGATGTGAG
5′ Ph-TATNGCAGGTACGTCGTACC GCGGCCGC GCAACGCGCTATGCTCGTag
5′ Ph-TAGNGCAGGTACGTCGTACC GCGGCCGC GACTgTGGAACTACGACGATCg
5′ Ph-TACNGCAGGTACGTCGTACC GCGGCCGC aGCaGACCGAACCCTAGTCGC
5′ Ph-TAANGCAGGTACGTCGTACC GCGGCCGC cATACGTCGTAgggTTCGCGA

5′ Ph-GTTNGCAGGTACGTCGTACC GCGGCCGC ctCTCATACGCGTCTGCGCGT
5′ Ph-GTGNGCAGGTACGTCGTACC GCGGCCGC gAGTgTGCCTTACGTCGAGttc
5′ Ph-GTCNGCAGGTACGTCGTACC GCGGCCGC GTcACGTtGCGGCCTTAGTC
5′ Ph-GTANGCAGGTACGTCGTACC GCGGCCGC GagGTACGAgACTTGACACACG
5′ Ph-GGTNGCAGGTACGTCGTACC GCGGCCGC GACcAGttGCCTAACGGAcACT
5′ Ph-GGGNGCAGGTACGTCGTACC GCGGCCGC GCAACTAGTCTCGACCTGCGA
5′ Ph-GGCNGCAGGTACGTCGTACC GCGGCCGC GTACCTCGACGACCGTACTGTg
5′ Ph-GGANGCAGGTACGTCGTACC GCGGCCGC ACGCGTGATAGTACGGAGTCG

5′ Ph-GCTNGCAGGTACGTCGTACC GCGGCCGC CACTAGAGCGGCGTCAGTCTA
5′ Ph-GCGNGCAGGTACGTCGTACC GCGGCCGC GCACAGCGCTAGCACAGGA
5′ Ph-GCCNGCAGGTACGTCGTACC GCGGCCGC TACCGACAGTCCTCTGCGTGC
5′ Ph-GCANGCAGGTACGTCGTACC GCGGCCGC CTACGCTACGTTGCGAAGAAGGTA
5′ Ph-GATNGCAGGTACGTCGTACC GCGGCCGC GTCTGTCGTACCTGTCAGTGACTg
5′ Ph-GAGNGCAGGTACGTCGTACC GCGGCCGC ATCGAACCGTGCTCCTTGG
5′ Ph-GACNGCAGGTACGTCGTACC GCGGCCGC AGGTTGAGGTGTACGCGATAGC
5′ Ph-GAANGCAGGTACGTCGTACC GCGGCCGC GACTTcAACCCCTGACGTACACa

5′ Ph-CTTNGCAGGTACGTCGTACC GCGGCCGC CTACTCGCGAGAGAGGGCTATG
5′ Ph-CTGNGCAGGTACGTCGTACC GCGGCCGC CTTGATCCGTAGTCGAGACGG
5′ Ph-CTCNGCAGGTACGTCGTACC GCGGCCGC GTACAGACGTAGCGATCGCaG
5′ Ph-CTANGCAGGTACGTCGTACC GCGGCCGC gTGACTAACGAGGTCTGTAAGCTa
5′ Ph-CGTNGCAGGTACGTCGTACC GCGGCCGC GTCTgAGAGTCGACTgCGCTAAG
5′ Ph-CGGNGCAGGTACGTCGTACC GCGGCCGC CTcAGTAAGCCGGAGTCTAGCTAg
5′ Ph-CGCNGCAGGTACGTCGTACC GCGGCCGC CGCCCTAAACGGGATCGAGCGA
5′ Ph-CGANGCAGGTACGTCGTACC GCGGCCGC CGTACAGGCTAGGGGTTAGTCG

FIG. 20